Towards Large-Scale Computational Science and Engineering with Quantifiable Uncertainty

Tan Bui-Thanh

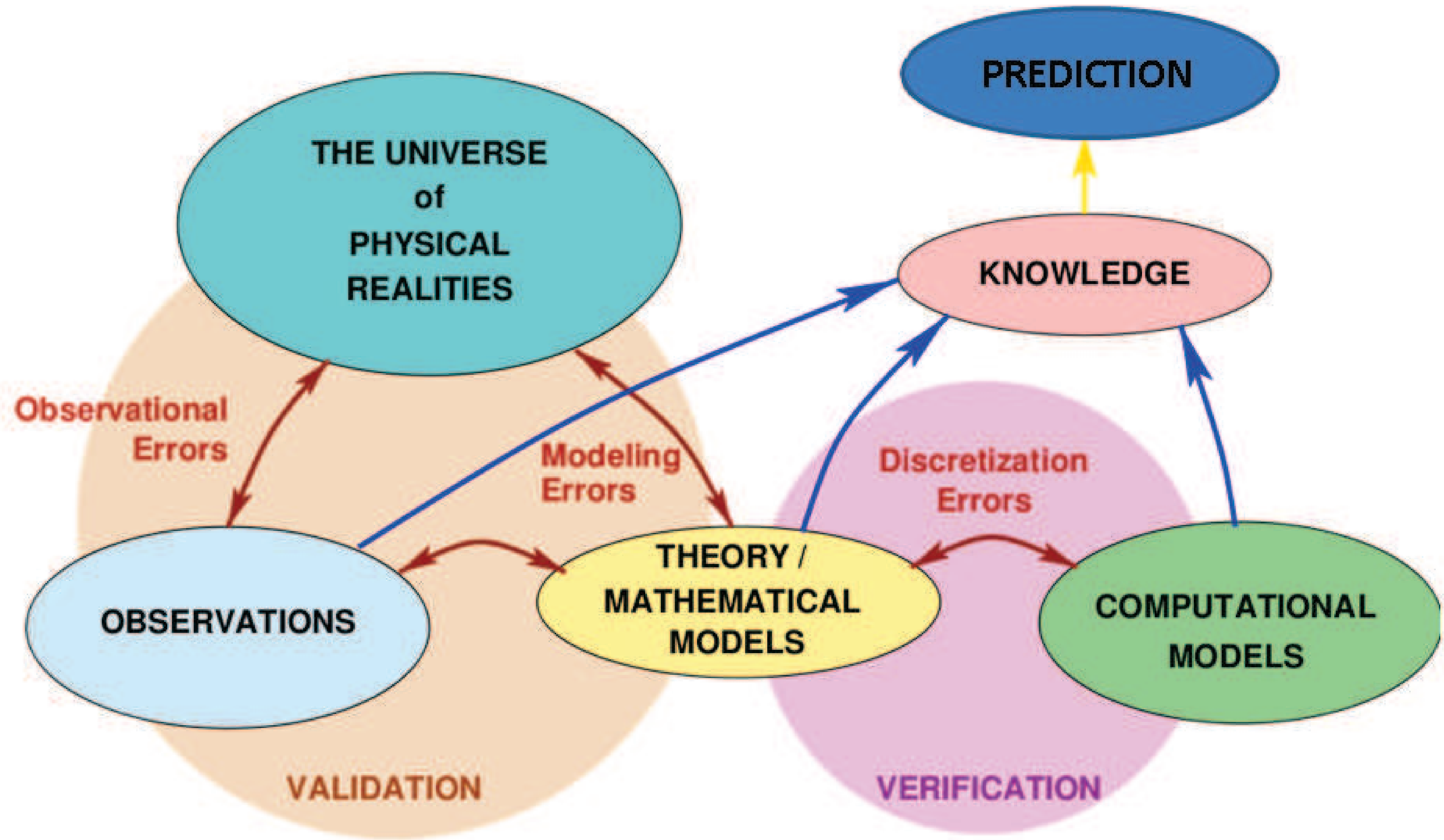
Department of Aerospace Engineering and Engineering Mechanics
Institute for Computational Engineering and Sciences (ICES)
The University of Texas at Austin

Joint work with Carsten Burstedde, Omar Ghattas, James R. Martin, Georg Stadler, and Lucas Wilcox

UQ Winter School
Geilo, Norway Jan 18-22, 2015

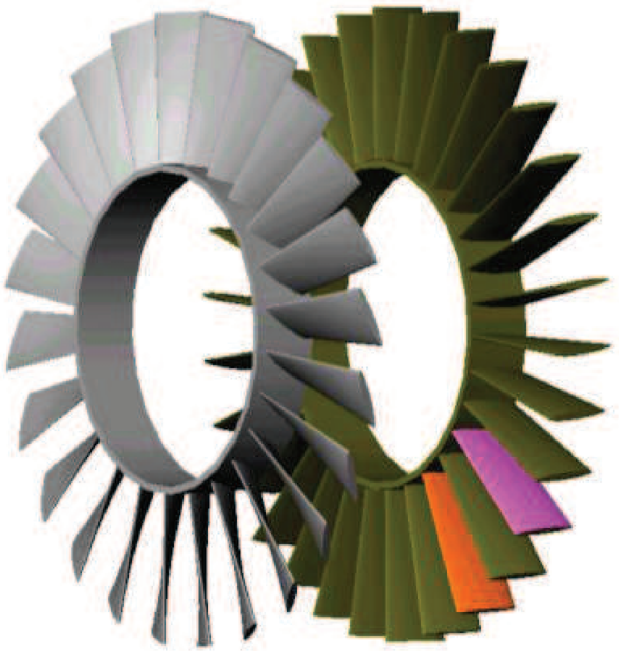
Imperfect path to knowledge

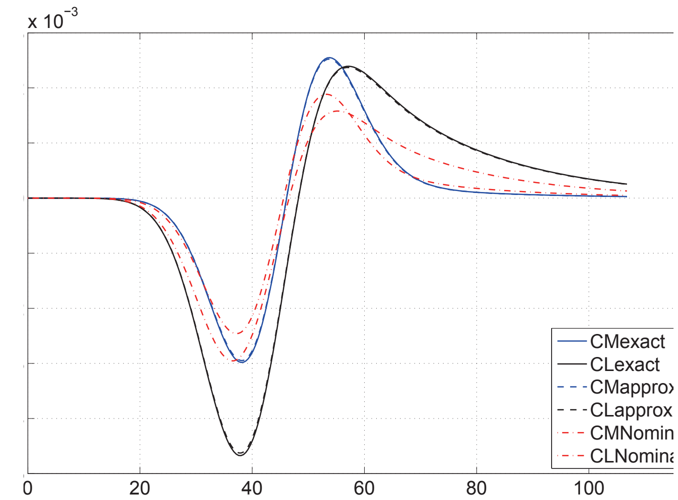
Courtesy of Oden *et al.*



Large-scale computation under uncertainty

CFD Probabilistic analysis for mistuned bladed disk





Geometric mistuning

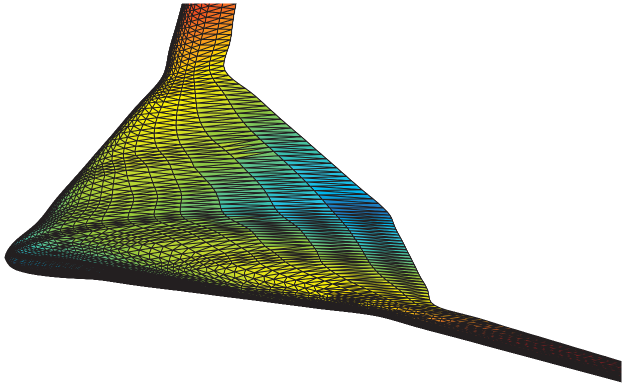
Random blade-to-blade variation due to imperfection of manufacturing process $\Rightarrow$ Large impact on forced response $\Rightarrow$ high-cycle fatigue properties of engine

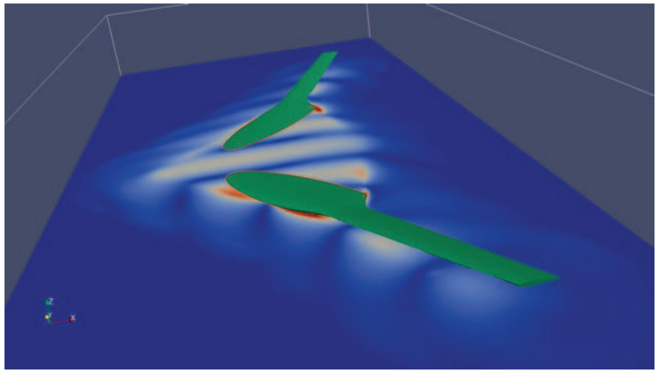
Challenge

How to propagate uncertainty in blade geometries to a quantity of interest via large-scale CFD simulation with $\mathcal{O}\left(10^6\right)$ degree of freedoms?

Large-scale computation under uncertainty

Inverse electromagnetic scattering





Randomness

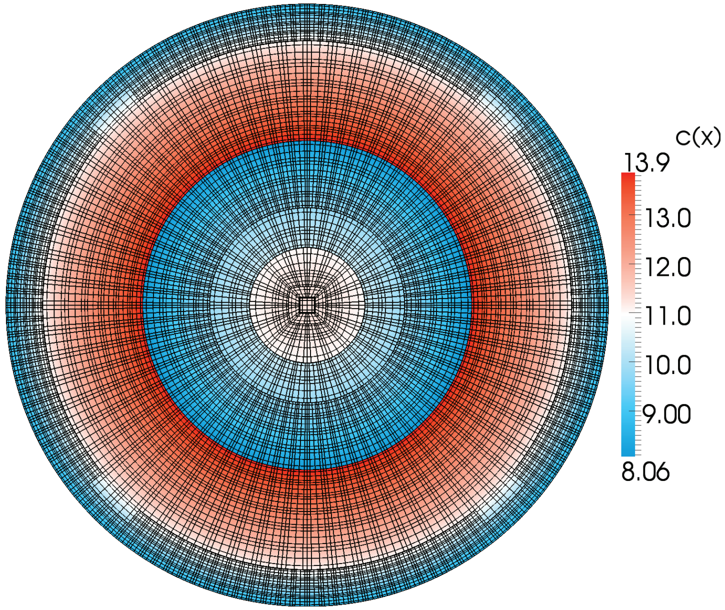
- Random errors in measurements are unavoidable
- Inadequacy of the mathematical model (Maxwell equations)

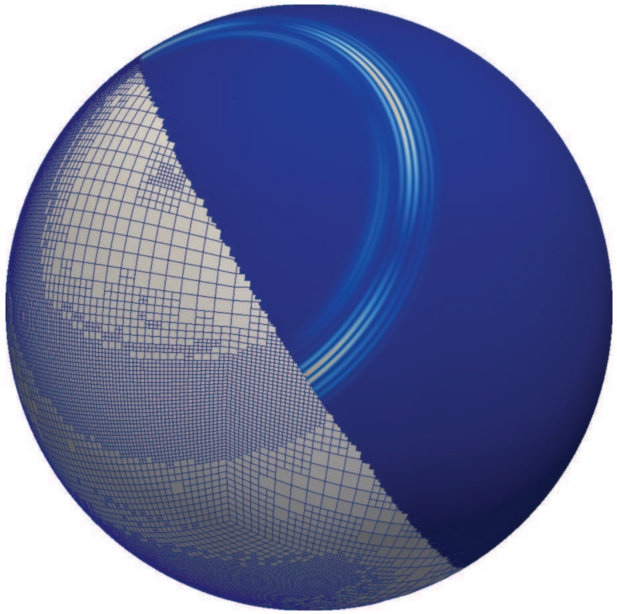
Challenge

How to invert for the invisible shape/medium using computational electromagnetics with $\mathcal{O}\left(10^6\right)$ degree of freedoms?

Large-scale computation under uncertainty

Full wave form seismic inversion





Randomness

- Random errors in seismometer measurements are unavoidable
- Inadequacy of the mathematical model (elastodynamics)

Challenge

How to image the earth interior using forward computational model with with $\mathcal{O}\left(10^9\right)$ degree of freedoms?

Tentative Plan for the Lectures

Contents

1. From basic Bayesian probability theory (lecture-oriented)
   - Probability space, random variables, distribution
   - Conditional probability, Bayes' formula, prior, likelihood, posterior
   - Construction of likelihood and prior
   - Relation between Bayesian inverse and deterministic inverse problems
   - Monte Carlo and classical limit theorems

Tentative Plan for the Lectures

Contents

1. From basic Bayesian probability theory (lecture-oriented)
   - Probability space, random variables, distribution
   - Conditional probability, Bayes' formula, prior, likelihood, posterior
   - Construction of likelihood and prior
   - Relation between Bayesian inverse and deterministic inverse problems
   - Monte Carlo and classical limit theorems
2. Advanced topics (lecture-oriented)
   - Advanced MCMC techiniques (Langevin, stochastic Newton, Hamiltonian, randomized MAP)
   - Bayesian inversion in infinite dimensions
   - Discretization-invariant MCMC methods

Tentative Plan for the Lectures

Contents

1. From basic Bayesian probability theory (lecture-oriented)
   - Probability space, random variables, distribution
   - Conditional probability, Bayes' formula, prior, likelihood, posterior
   - Construction of likelihood and prior
   - Relation between Bayesian inverse and deterministic inverse problems
   - Monte Carlo and classical limit theorems
2. Advanced topics (lecture-oriented)
   - Advanced MCMC techiniques (Langevin, stochastic Newton, Hamiltonian, randomized MAP)
   - Bayesian inversion in infinite dimensions
   - Discretization-invariant MCMC methods
3. Large-scale Bayesian inverse problems (seminar-oriented)
   - Reduce-then-sample approach
   - Sample-then-reduce approach
   - Compactness, convergence of discrete Bayesian approaches, etc
   - Applications to large-scale inverse problems
   - Big-data in large-scale inverse problems

Outline

1. Doubly Infinite Dimensional Problems for UQ
2. Forward propagation of uncertainty in CFD (Reduce-then-Sample)
3. Statistical inverse problem in electromagnetics (Reduced-then-Sample)
4. Ultra-scale Seismic Wave Inversion (Sample-then-Reduce)
   - Infinite Dimensional Bayesian inference
   - Infinite Dimensional Bayesian inference: Derivation of $A^*\Gamma^{-1}A$
   - Infinite Dimensional Bayesian Inference: Compactness of $A^*\Gamma^{-1}A$
   - Scalable Discontinuous Galerkin for seismic waves: Scalability of A
5. Summary

Large-scale uncertainty quantification in high dimensions

Common challenge (**Doubly infinite dimensional problem**)

- Curse of dimensionality
- Need to repeatedly solve the large-scale computational system under consideration

Large-scale uncertainty quantification in high dimensions

Common challenge (**Doubly infinite dimensional problem**)

- Curse of dimensionality
- Need to repeatedly solve the large-scale computational system under consideration

Solution 1: Reduce-then-sample

Construct

- reduced basis (reduced-order) model or
- an accurate surrogate model

that is inexpensive to solve

Large-scale uncertainty quantification in high dimensions

Common challenge (**Doubly infinite dimensional problem**)

- Curse of dimensionality
- Need to repeatedly solve the large-scale computational system under consideration

Solution 1: Reduce-then-sample

Construct

- reduced basis (reduced-order) model or
- an accurate surrogate model

that is inexpensive to solve

Solution 2: Sample-then-reduce

Work directly with high-fidelity model but only explore important subspaces/directions

Outline

Forced Response Blade Example

2D Euler

$$\frac{\partial}{\partial t}\begin{pmatrix}\rho\\ \rho u\\ \rho v\\ \rho e\end{pmatrix}+\frac{\partial}{\partial x}\begin{pmatrix}\rho u\\ \rho u^2+p\\ \rho uv\\ u\left(\rho e+p\right)\end{pmatrix}+\frac{\partial}{\partial y}\begin{pmatrix}\rho v\\ \rho uv\\ \rho v^2+p\\ v\left(\rho e+p\right)\end{pmatrix}=0$$

- Discontinuous Galerkin discretization

Static pressure

Forced Response Blade Example

Static pressure

2D Euler

$$\frac{\partial}{\partial t}\begin{pmatrix}\rho\\ \rho u\\ \rho v\\ \rho e\end{pmatrix}+\frac{\partial}{\partial x}\begin{pmatrix}\rho u\\ \rho u^2+p\\ \rho uv\\ u\left(\rho e+p\right)\end{pmatrix}+\frac{\partial}{\partial y}\begin{pmatrix}\rho v\\ \rho uv\\ \rho v^2+p\\ v\left(\rho e+p\right)\end{pmatrix}=0$$

- Discontinuous Galerkin discretization

Full model

$$E\left(\mathbf{u}\right)\dot{x}=A\left(\mathbf{u}\right)x+B\left(\mathbf{u}\right)z$$
$$y=C\left(\mathbf{u}\right)x$$

Reduced basis approach

$x=\Phi x_r$

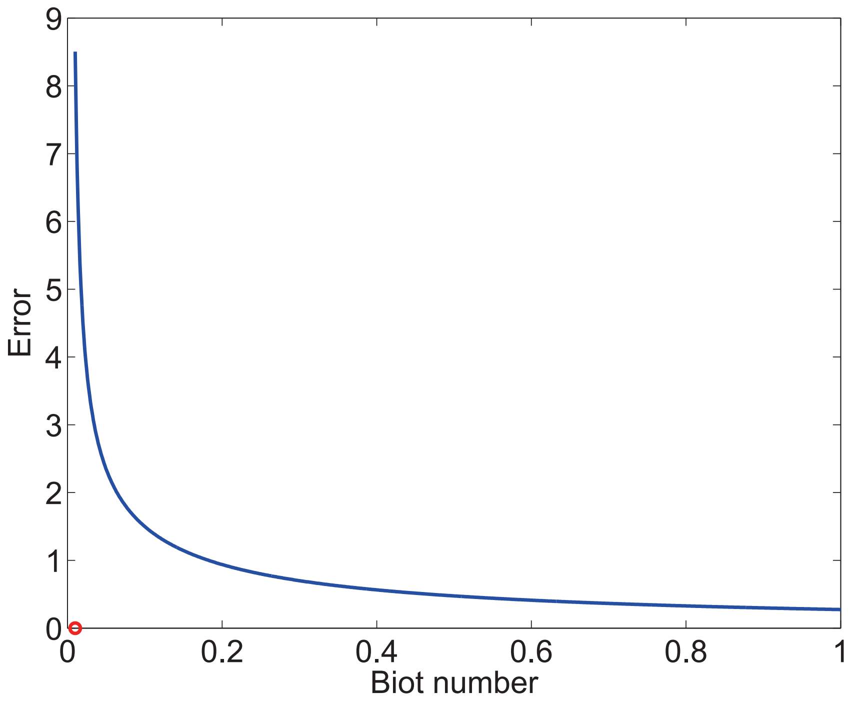
$$\Psi^T E\Phi\dot{x}_r=\Psi^T A\Phi x_r+\Psi^T Bu$$
$$y_r=C\Phi x_r$$

Petrov-Galerkin projection

Greedy optimization approach to find the reduced bases Φ, Ψ

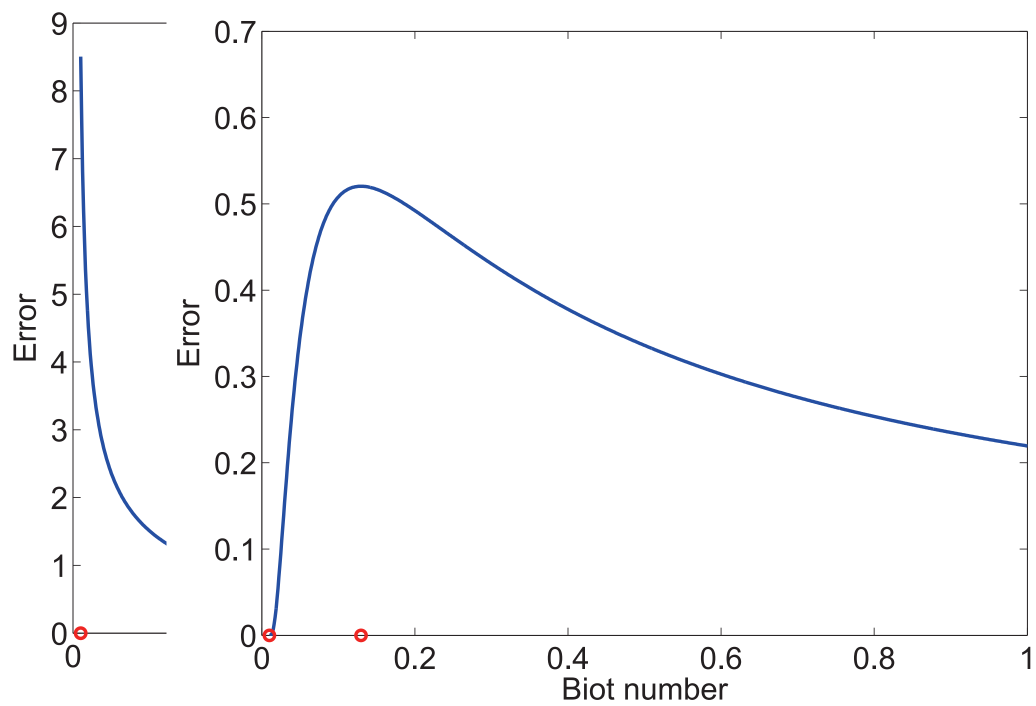
The Model-Constrained Adaptive Sampling Algorithm

1. Given a reduced basis Φ and initial guess $\mathbf{u}^0$. Find $\mathbf{u}^* = \arg\max_{\mathbf{u}} \|y(\mathbf{u}) - y_r(\mathbf{u})\|_2^2$
2. If $\|y(\mathbf{u}^*) - y_r(\mathbf{u}^*)\|_2^2 \leq \epsilon$, then terminate the algorithm. If not, go to the next step.
3. With $\mathbf{u} = \mathbf{u}^*$, solve the full system to compute the state solutions $x(\mathbf{u}^*)$, which is then used to update the basis Φ. Go to step 1.



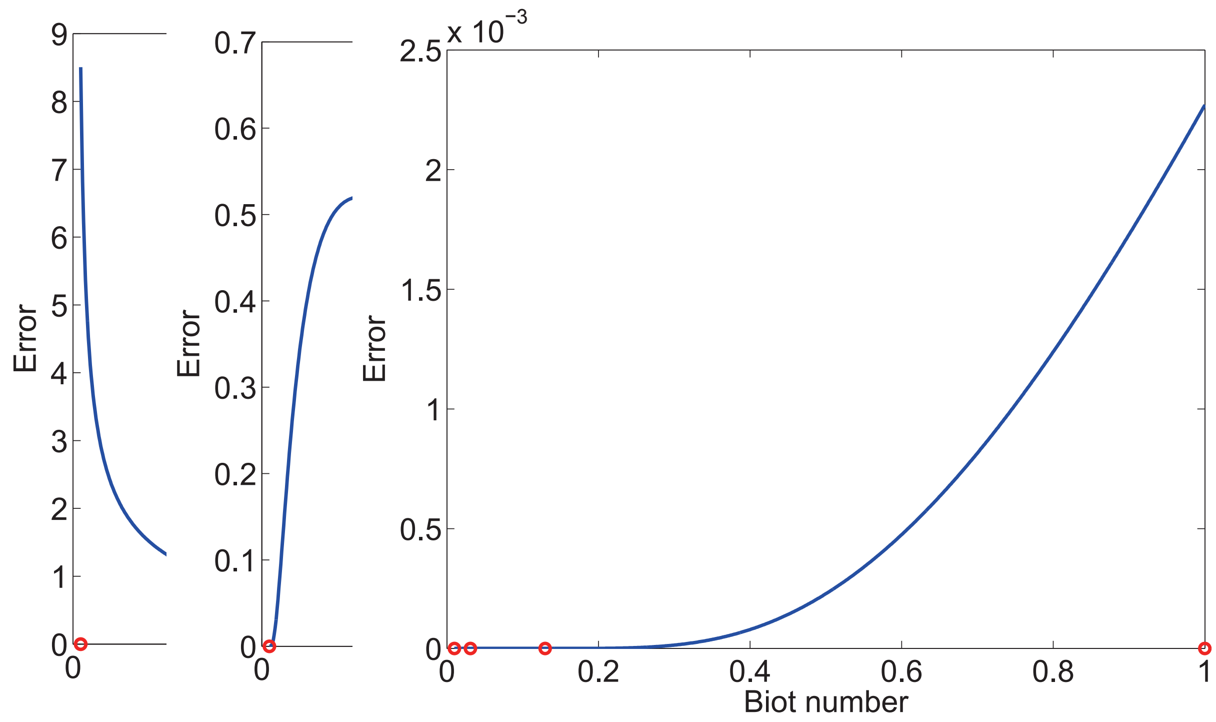
The Model-Constrained Adaptive Sampling Algorithm

1. Given a reduced basis Φ and initial guess $\mathbf{u}^0$. Find $\mathbf{u}^* = \arg\max_{\mathbf{u}} \|y(\mathbf{u}) - y_r(\mathbf{u})\|_2^2$
2. If $\|y(\mathbf{u}^*) - y_r(\mathbf{u}^*)\|_2^2 \leq \epsilon$, then terminate the algorithm. If not, go to the next step.
3. With $\mathbf{u} = \mathbf{u}^*$, solve the full system to compute the state solutions $x(\mathbf{u}^*)$, which is then used to update the basis Φ. Go to step 1.



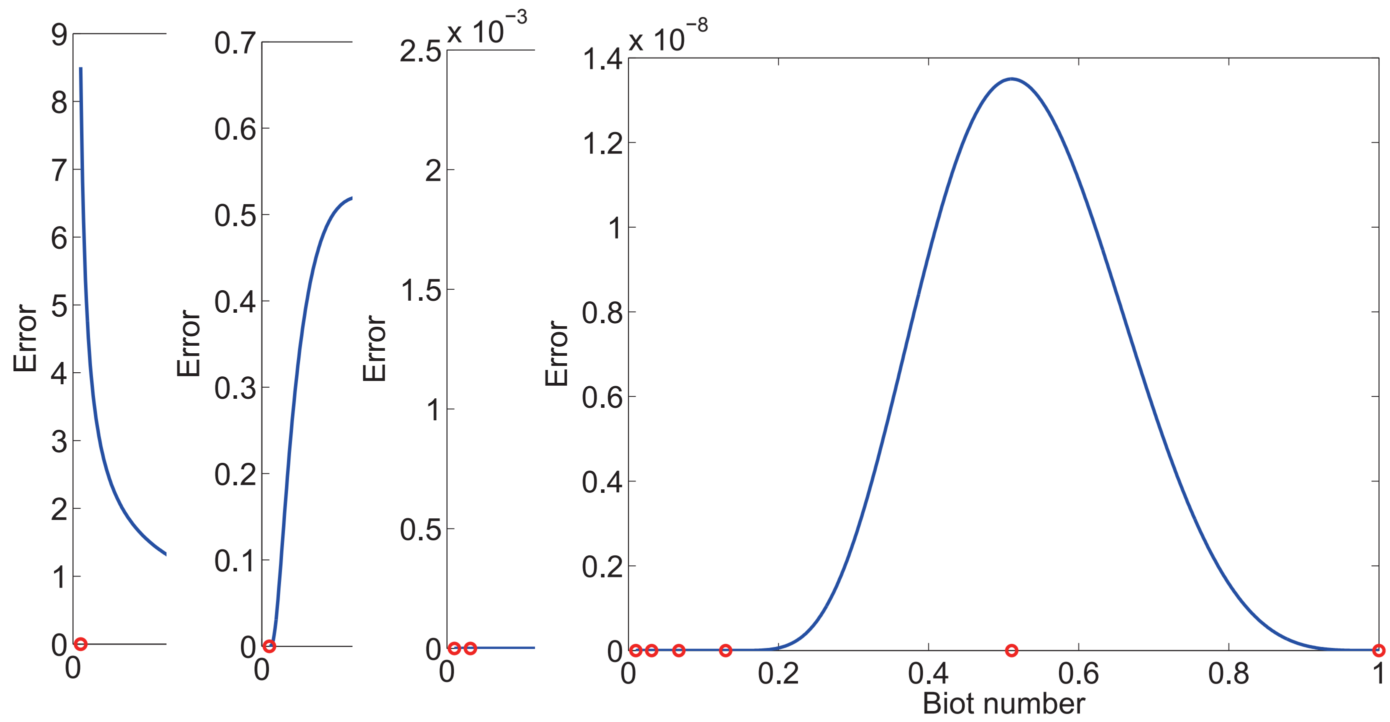
The Model-Constrained Adaptive Sampling Algorithm

1. Given a reduced basis Φ and initial guess $\mathbf{u}^0$. Find $\mathbf{u}^* = \arg\max_{\mathbf{u}} \|y(\mathbf{u}) - y_r(\mathbf{u})\|_2^2$
2. If $\|y(\mathbf{u}^*) - y_r(\mathbf{u}^*)\|_2^2 \leq \epsilon$, then terminate the algorithm. If not, go to the next step.
3. With $\mathbf{u} = \mathbf{u}^*$, solve the full system to compute the state solutions $x(\mathbf{u}^*)$, which is then used to update the basis Φ. Go to step 1.



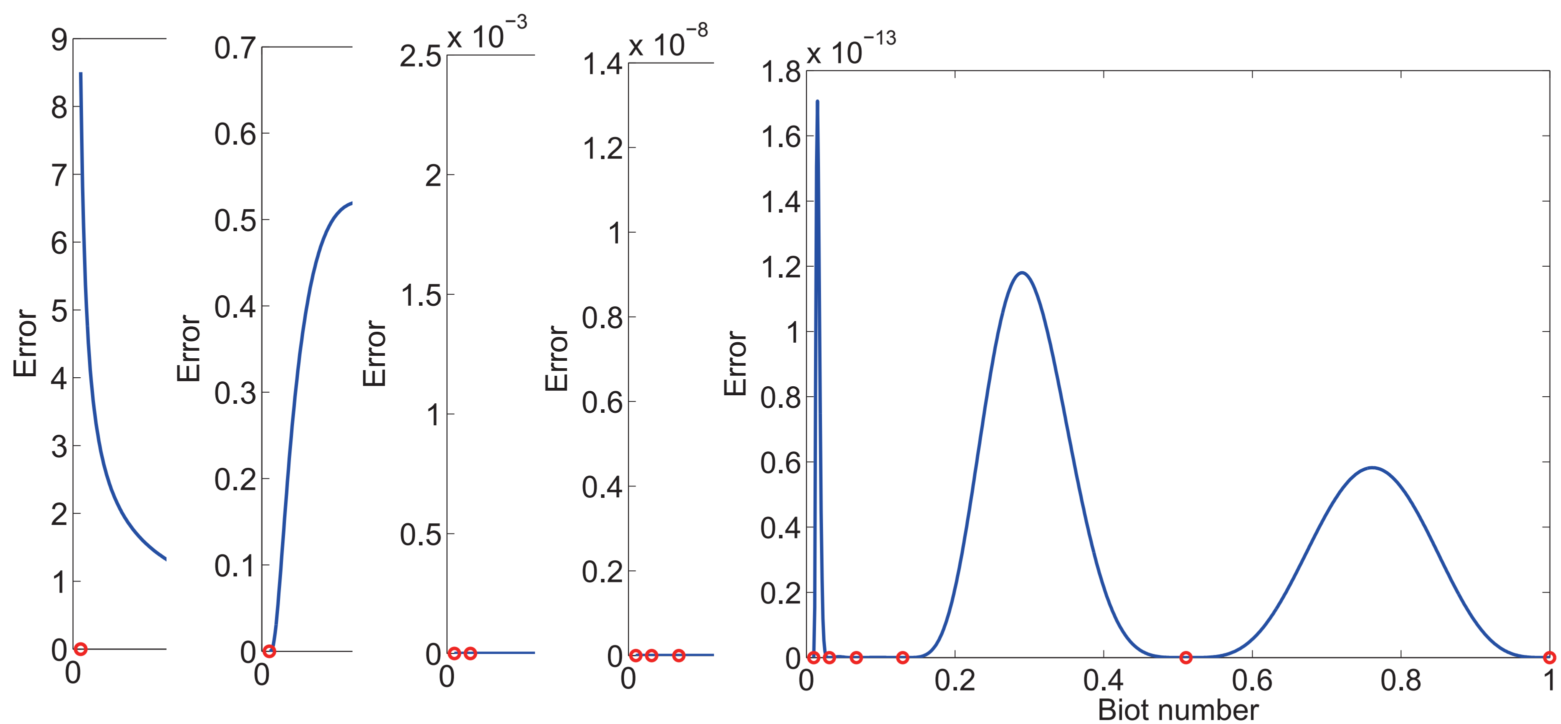
The Model-Constrained Adaptive Sampling Algorithm

1. Given a reduced basis Φ and initial guess $\mathbf{u}^0$. Find $\mathbf{u}^* = \arg\max_{\mathbf{u}} \|y(\mathbf{u}) - y_r(\mathbf{u})\|_2^2$
2. If $\|y(\mathbf{u}^*) - y_r(\mathbf{u}^*)\|_2^2 \leq \epsilon$, then terminate the algorithm. If not, go to the next step.
3. With $\mathbf{u} = \mathbf{u}^*$, solve the full system to compute the state solutions $x(\mathbf{u}^*)$, which is then used to update the basis Φ. Go to step 1.



The Model-Constrained Adaptive Sampling Algorithm

1. Given a reduced basis Φ and initial guess $\mathbf{u}^0$. Find $\mathbf{u}^* = \arg\max_{\mathbf{u}} \|y(\mathbf{u}) - y_r(\mathbf{u})\|_2^2$
2. If $\|y(\mathbf{u}^*) - y_r(\mathbf{u}^*)\|_2^2 \leq \epsilon$, then terminate the algorithm. If not, go to the next step.
3. With $\mathbf{u} = \mathbf{u}^*$, solve the full system to compute the state solutions $x(\mathbf{u}^*)$, which is then used to update the basis Φ. Go to step 1.



Well-definedness of the Adaptive Sampling Method

Theorem

The proposed adaptive sampling algorithm is well-defined in the sense that:

1. *it terminates in finite time and*
2. *all the sampled points are distinct.*

Details in:

- Bui-Thanh, T., Willcox, K., and Ghattas, O., *Model Reduction for Large-Scale Systems with High-Dimensional Parametric Input Space*, **SIAM Journal on Scientific Computing**, 30(6), pp. 3270–3288, 2008.
- Bui-Thanh, T., Willcox, K., and Ghattas, O., *Parametric Reduced-Order Models for Probabilistic Analysis of Unsteady Aerodynamic Applications*, **AIAA Journal**, 46(10), pp. 2520–2529, 2008.
- Bui-Thanh, T., *Model-Constrained Optimization Methods for Reduction of Parameterized Large-Scale Systems.* **MIT**, PhD thesis, 2007.

Forced Response Blade Example (Monte Carlo Simulation)

- Work per cycle (WPC) of two blades moving 180^o degrees out of phase
- Geometry described by 4 parameters
- Same 10,000 random samples for both full and reduced models

	Full CFD	Reduced CFD
Model size	103,008	201
Number of nonzeros	2,846,056	40,401
Offline cost	—	2.8 hours
Online cost	501.1 hours	0.21 hours
Blade 1 WPC mean	-1.8572	-1.8573
Blade 1 WPC variance	2.687e-4	2.6819e-4
Blade 2 WPC mean	-1.8581	-1.8580
Blade 2 WPC variance	2.797e-4	2.799e-4

Details in: Bui-Thanh, T., Willcox, K., and Ghattas, O., *Model Reduction for Large-Scale Systems with High-Dimensional Parametric Input Space*, **SIAM Journal on Scientific Computing**, 30(6), pp. 3270–3288, 2008.

Forced Response Blade Example (Monte Carlo Simulation)

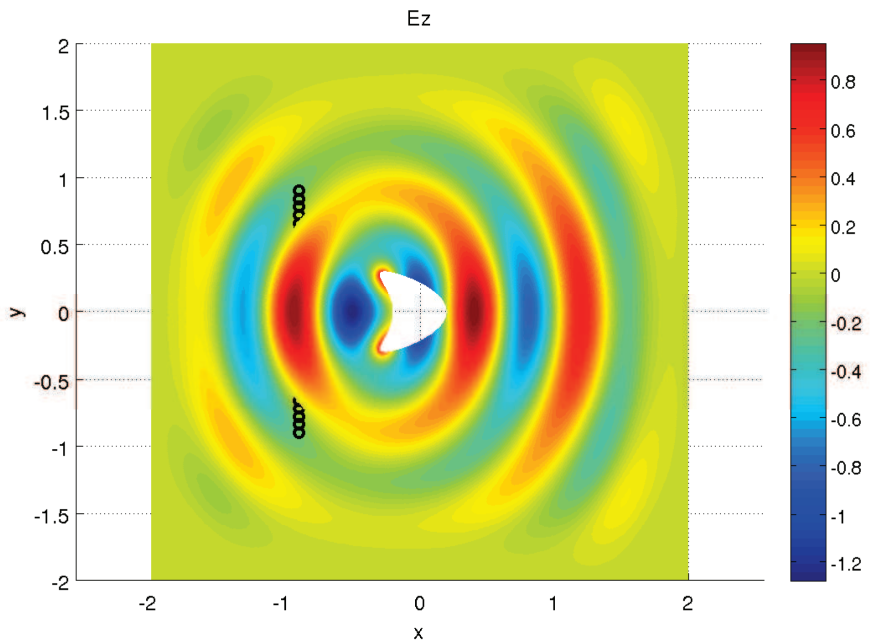
Outline

Inverse Shape Electromagnetic Scattering Problem

Maxwell Equations:

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \qquad \text{(Faraday)}$$

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t}, \qquad \text{(Ampere)}$$



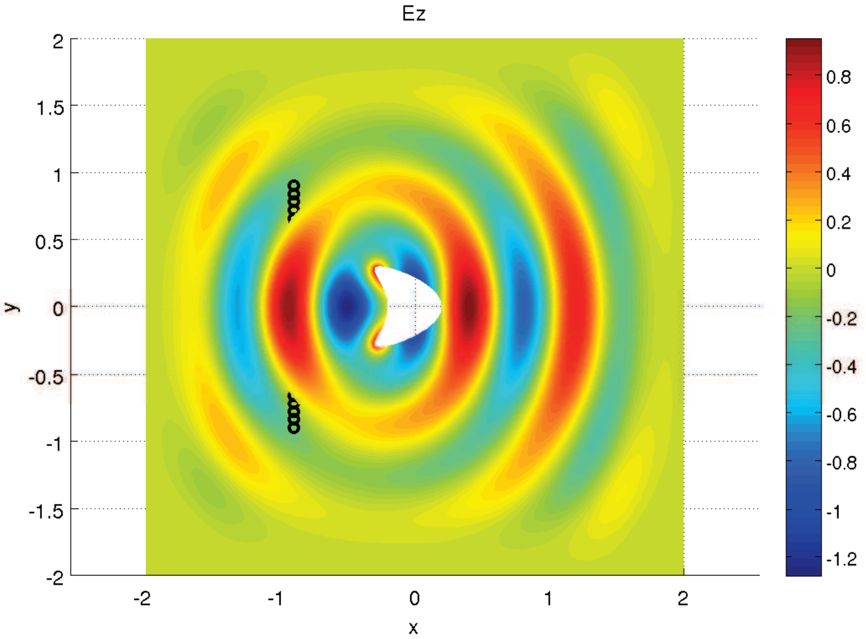
$\mathbf{E}$: Electric field, $\mathbf{H}$: Magnetic field, μ: permeability, ϵ: permittivity

Inverse Shape Electromagnetic Scattering Problem

Maxwell Equations:

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \qquad \text{(Faraday)}$$

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t}, \qquad \text{(Ampere)}$$



$\mathbf{E}$: Electric field, $\mathbf{H}$: Magnetic field, μ: permeability, ϵ: permittivity

Forward problem (discontinuous Galerkin discretization)

$$\mathbf{y} = \mathbf{G}(\mathbf{u})$$

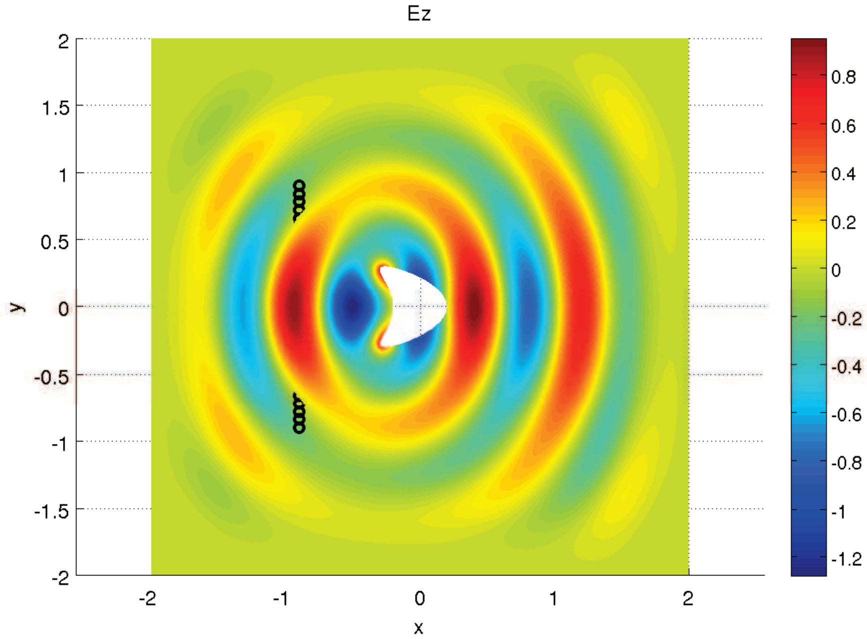
where $\mathbf{G}$ maps shape parameters $\mathbf{u}$ to electric/magnetic field $\mathbf{y}$ at the measurement points

Inverse Shape Electromagnetic Scattering Problem

Maxwell Equations:

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \qquad \text{(Faraday)}$$

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t}, \qquad \text{(Ampere)}$$



$\mathbf{E}$: Electric field, $\mathbf{H}$: Magnetic field, μ: permeability, ϵ: permittivity

Forward problem (discontinuous Galerkin discretization)

$$\mathbf{y} = \mathbf{G}(\mathbf{u})$$

where $\mathbf{G}$ maps shape parameters $\mathbf{u}$ to electric/magnetic field $\mathbf{y}$ at the measurement points

Inverse Problem

Given (possibly noise-corrupted) measurements on $\mathbf{y}$, infer $\mathbf{u}$?

The Bayesian Statistical Inversion Framework

Bayes Theorem

Solution to the inverse problem is given as a posterior PDF over parameter space:

$$\pi_{\mathrm{post}}(\mathbf{u}|\boldsymbol{y}_{\mathrm{obs}}) \propto \pi_{\mathrm{pr}}(\mathbf{u})\pi_{\mathrm{like}}(\boldsymbol{y}_{\mathrm{obs}}|\mathbf{u})$$

The Bayesian Statistical Inversion Framework

Bayes Theorem

Solution to the inverse problem is given as a posterior PDF over parameter space:

$$\pi_{\text{post}}(\mathbf{u}|\boldsymbol{y}_{\text{obs}}) \propto \pi_{\text{pr}}(\mathbf{u})\pi_{\text{like}}(\boldsymbol{y}_{\text{obs}}|\mathbf{u})$$

Prior knowledge: The obstacle is smooth:

$$\pi_{\text{pr}}(\mathbf{u}) \propto \exp\left(-\lambda \int_0^{2\pi} r''(\mathbf{u}) d\theta\right)$$

The Bayesian Statistical Inversion Framework

Bayes Theorem

Solution to the inverse problem is given as a posterior PDF over parameter space:

$$\pi_{\text{post}}(\mathbf{u}|\boldsymbol{y}_{\text{obs}}) \propto \pi_{\text{pr}}(\mathbf{u})\pi_{\text{like}}(\boldsymbol{y}_{\text{obs}}|\mathbf{u})$$

Prior knowledge: The obstacle is smooth:

$$\pi_{\text{pr}}(\mathbf{u}) \propto \exp\left(-\lambda \int_0^{2\pi} r''(\mathbf{u})d\theta\right)$$

Likelihood: Additive Gaussian noise, for example,

$$\pi_{\text{like}}(\boldsymbol{y}_{\text{obs}}|\mathbf{u}) \propto \exp\Big(-\frac{1}{2}(\boldsymbol{G}(\mathbf{u}) - \boldsymbol{y}_{\text{obs}})^T\Sigma_{\text{noise}}^{-1}(\boldsymbol{G}(\mathbf{u}) - \boldsymbol{y}_{\text{obs}})\Big)$$

Challenge: Expensive Forward Solve

$$\pi_{\text{post}}(\mathbf{u}|\boldsymbol{y}_{\text{obs}}) \propto \pi_{\text{pr}}(\mathbf{u}) \times \underbrace{\pi_{\text{like}}(\boldsymbol{y}_{\text{obs}}|\mathbf{u})}_{\text{Computationally expensive forward model:}\mathbf{y}=\mathbf{G}(\mathbf{u})}$$

Challenge: Expensive Forward Solve

$$\pi_{\text{post}}(\mathbf{u}|\boldsymbol{y}_{\text{obs}}) \propto \pi_{\text{pr}}(\mathbf{u}) \times \underbrace{\pi_{\text{like}}(\boldsymbol{y}_{\text{obs}}|\mathbf{u})}_{\text{Computationally expensive forward model:}\mathbf{y}=\mathbf{G}(\mathbf{u})}$$

Approximate the likelihood

Reduced basis method, polynomial chaos, and etc

Challenge: Expensive Forward Solve

$$\pi_{\text{post}}(\mathbf{u}|\boldsymbol{y}_{\text{obs}}) \propto \pi_{\text{pr}}(\mathbf{u}) \times \underbrace{\pi_{\text{like}}(\boldsymbol{y}_{\text{obs}}|\mathbf{u})}_{\text{Computationally expensive forward model:}\mathbf{y}=\mathbf{G}(\mathbf{u})}$$

Approximate the likelihood

Reduced basis method, polynomial chaos, and etc

Approximate the posterior

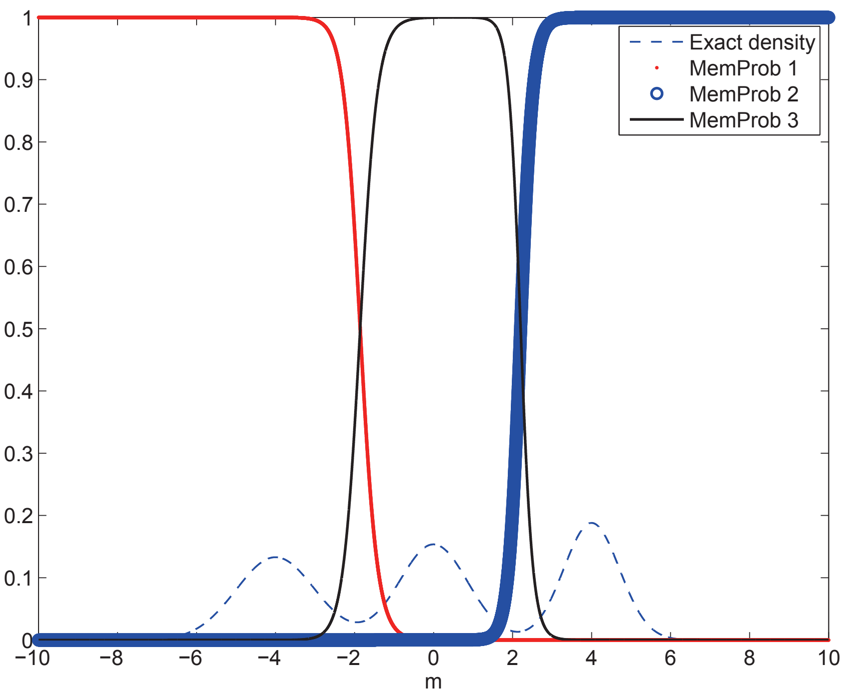
1. The posterior is nicer than the likelihood due to the prior contribution
2. The posterior is scalar function of parameters $\mathbf{u}$ only

We propose a **Hessian-based Adaptive Gaussian Process** response surface

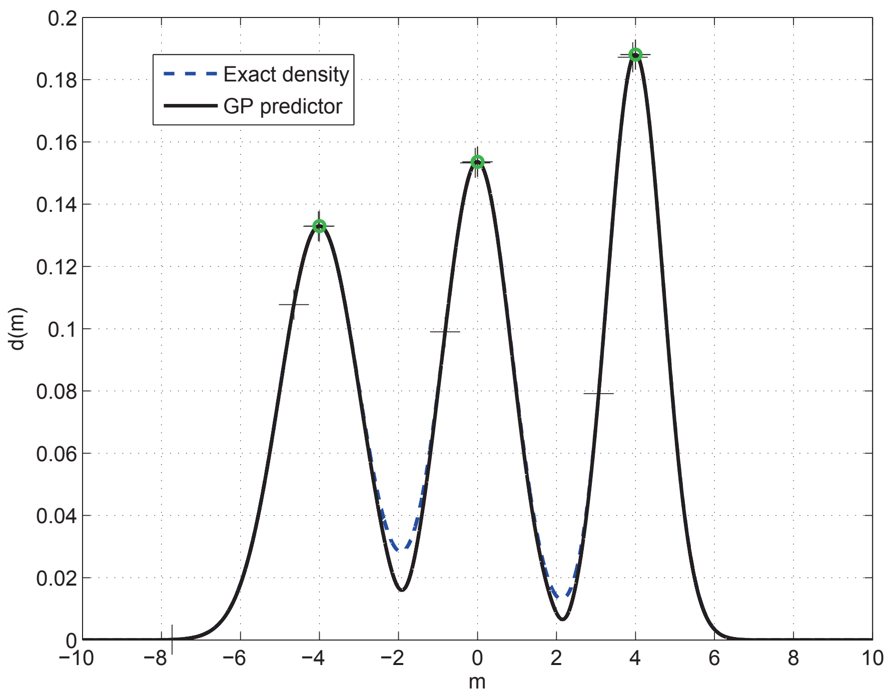
Hessian-based Adaptive Gaussian Process

Main idea (mitigating the curse of dimensionality)

1. Use Adaptive Sampling Algorithm to find the modes
2. Approximate the covariance matrix (Hessian inverse)
3. Partition parameter space using membership probabilities
4. Approximate the posterior with local Gaussian in subdomains
5. Glue all the local Gaussian approximations



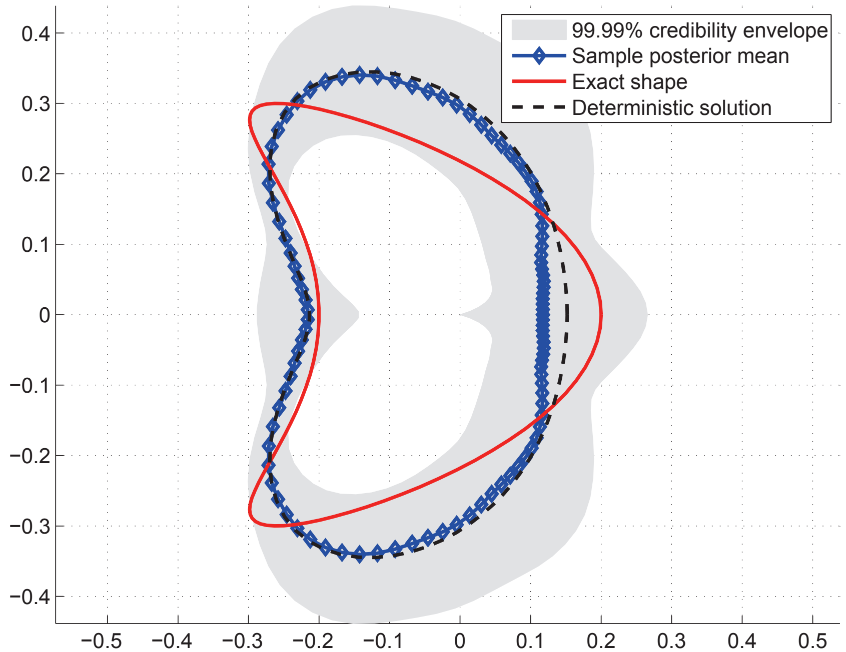
Membership probabilities



GP predictor versus exact

Inverse shape electromagnetic scattering

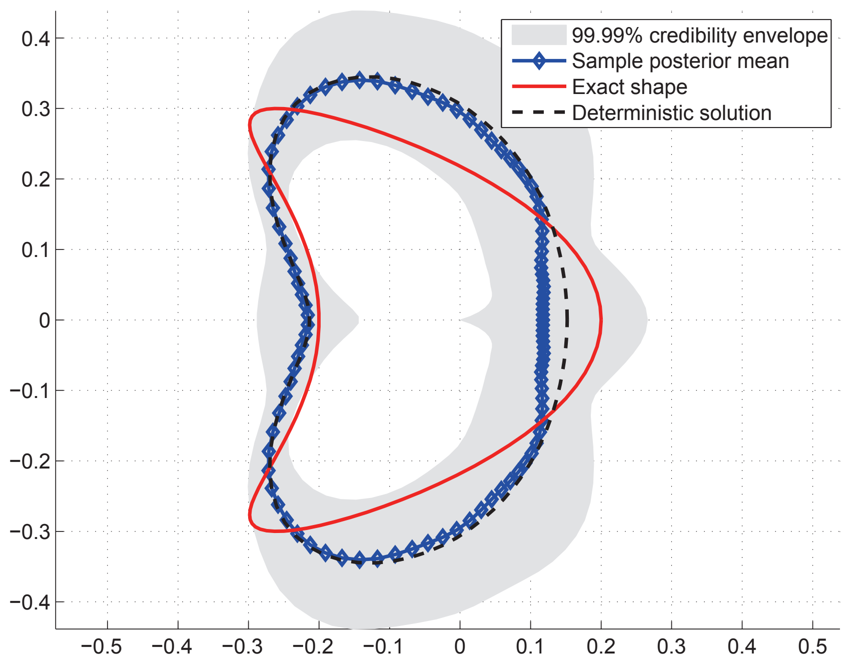
- discontinuous Galerkin discretization with 80,892 state variables
- 24 shape parameters
- 1 million MCMC simulations for the Gaussian process response surface



Details in: Bui-Thanh, T., Ghattas, O., and Higdon, D., *Adaptive Hessian-based Non-stationary Gaussian Process Response Surface Method for Probability Density Approximation with Application to Bayesian Solution of Large-scale Inverse Problems*, **SIAM Journal on Scientific Computing**, 34(6), pp. A2837–A2871, 2012.

Inverse shape electromagnetic scattering

- discontinuous Galerkin discretization with 80,892 state variables
- 24 shape parameters
- 1 million MCMC simulations for the Gaussian process response surface



	Offline time
Gaussian process	33 hours
Exact Posterior	0 hours

	Online time
Gaussian process	0.96 hours
Exact Posterior	8802.35 hours

Details in: Bui-Thanh, T., Ghattas, O., and Higdon, D., *Adaptive Hessian-based Non-stationary Gaussian Process Response Surface Method for Probability Density Approximation with Application to Bayesian Solution of Large-scale Inverse Problems*, **SIAM Journal on Scientific Computing**, 34(6), pp. A2837–A2871, 2012.

Outline

Full wave form seismic wave inversion

$$\frac{\partial \boldsymbol{E}}{\partial t} = \frac{1}{2}\left(\nabla \boldsymbol{v} + \nabla^T \boldsymbol{v}\right),$$
$$\rho \frac{\partial \boldsymbol{v}}{\partial t} = \nabla \cdot (\mathsf{C}\boldsymbol{E}) + \boldsymbol{f}$$

Strain-velocity formulation

Animated by Greg Abram

- I: fourth-order identity tensor,
- $\boldsymbol{I}$: second-order identity tensor,
- $\boldsymbol{f}$: external volumetric forces,
- C: four-order material tensor.

- $\boldsymbol{E}$: strain tensor,
- $\boldsymbol{v}$: velocity vector,
- ρ: density,
- $\boldsymbol{e}_i$: ith unit vector,

Inverse problem statement

- Earth surface velocity at given locations is recorded
- Infer the wave velocities $c_s = \sqrt{\mu/\rho}$ and $c_p = \sqrt{(\lambda + 2\mu)/\rho}$

Infinite dimensional Bayesian statistical inference

Why infinite dimensions?

- The original inverse problem is posed in infinite dimensional spaces
- Algorithms for infinite dimensional setting are most likely independent of the dimension of discretization
- Finite dimensional Bayes formula breaks down!

Our moral

Do as much as we can on infinite dimensional level, and discretization is the last step

Infinite dimensional Bayesian statistical inference

Bayes' theorem in infinite dimensions

A Bayes' theorem in infinite dimensional spaces (Stuart 2010)

$$\frac{d\mu}{d\mu_0}(\mathbf{u}) \propto \exp\left(-\Phi\left(\boldsymbol{y}^{\text{obs}}, \mathbf{u}\right)\right)$$

defines the Radon-Nikodym derivative of the posterior probability measure μ with respect to the prior measure μ_0.

- μ_0: prior probability measure
- μ: posterior probability measure
- $\Phi\left(\boldsymbol{y}^{\text{obs}}, \mathbf{u}\right)$: misfit functional
- $\mathbf{u}$: unknown parameter
- $\boldsymbol{y}^{\text{obs}}$: observation data

Infinite dimensional Bayesian statistical inference

Linearized Bayesian solution about the MAP

- Compute the MAP
- Linearize the forward map about the MAP $\boldsymbol{y}^{\text{obs}} = f_0 + A(\mathbf{u}) + \eta$

Posterior becomes a Gaussian measure

$$\mathbf{u}|\boldsymbol{y}^{\text{obs}} \sim \mu = \mathcal{N}(\mathbf{m}, \mathcal{C}),$$

posterior mean

$$\mathbf{m} = \mathbb{E}[\mathbf{u}] = \mathbf{u}_0 + C_0 A^* \left(\Gamma + A\mathcal{C}_0 A^*\right)^{-1} \left(\boldsymbol{y}^{\text{obs}} - f_0 - A\mathbf{u}_0\right)$$

posterior covariance operator

$$\mathcal{C} = \left(A^*\Gamma^{-1}A + \mathcal{C}_0^{-1}\right)^{-1}$$

Infinite dimensional Bayesian statistical inference

Linearized Bayesian solution: Low rank approximation

posterior covariance operator: A low rank approximation

$$
\begin{aligned}
\mathcal{C} &= \left(A^*\Gamma^{-1}A + \mathcal{C}_0^{-1}\right)^{-1} \\
&= \mathcal{C}_0^{1/2}\left(\mathcal{C}_0^{1/2}\boldsymbol{A}^*\boldsymbol{\Gamma}^{-1}\boldsymbol{A}\,\mathcal{C}_0^{1/2} + \boldsymbol{I}\right)^{-1}\mathcal{C}_0^{1/2} \\
&\approx \mathcal{C}_0^{1/2}\left(\boldsymbol{V}_r\boldsymbol{\Lambda}_r\boldsymbol{V}_r^* + \boldsymbol{I}\right)^{-1}\mathcal{C}_0^{1/2} \\
&= \mathcal{C}_0 - \mathcal{C}_0^{1/2}\boldsymbol{V}_r\boldsymbol{D}_r\boldsymbol{V}_r^*\mathcal{C}_0^{1/2}
\end{aligned}
$$

- Low rank approximation only involves incremetal forward and incremental adjoint solve
- Then use Sherman-Morrison-Woodbury
- Relative to the prior uncertainty, the posterior uncertainty is reduced when observations are made.

Computation of action of Hessian in given direction

- Action of the Hessian operator in direction $\tilde{\mathsf{C}}$ at a point C given by

$$\mathcal{H}(\mathsf{C})\tilde{\mathsf{C}} := \int_0^T \left[\frac{1}{2}(\nabla\tilde{\boldsymbol{w}} + \nabla\tilde{\boldsymbol{w}}^T) \otimes \boldsymbol{E} + \frac{1}{2}(\nabla\boldsymbol{w} + \nabla\boldsymbol{w}^T) \otimes \tilde{\boldsymbol{E}}\right] dt + \mathcal{R}''(\mathsf{C})\tilde{\mathsf{C}}$$

- where $\tilde{\boldsymbol{v}}, \tilde{\boldsymbol{E}}$ satisfy the *incremental forward wave propagation equations*

$$\begin{aligned}
\rho\frac{\partial\tilde{\boldsymbol{v}}}{\partial t} - \nabla\cdot(\mathsf{C}\tilde{\boldsymbol{E}}) &= \nabla\cdot(\tilde{\mathsf{C}}\boldsymbol{E}) && \text{in } \Omega\times(0,T)\\
-\mathsf{C}\frac{\partial\tilde{\boldsymbol{E}}}{\partial t} + \frac{1}{2}\mathsf{C}(\nabla\tilde{\boldsymbol{v}} + \nabla\tilde{\boldsymbol{v}}^T) &= \boldsymbol{0} && \text{in } \Omega\times(0,T)\\
\rho\tilde{\boldsymbol{v}} = \mathsf{C}\tilde{\boldsymbol{E}} &= \boldsymbol{0} && \text{in } \Omega\times\{t=0\}\\
\mathsf{C}\tilde{\boldsymbol{E}}\boldsymbol{n} &= -\tilde{\mathsf{C}}\boldsymbol{E}\boldsymbol{n} && \text{on } \Gamma\times(0,T)
\end{aligned}$$

- and $\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{D}}$ satisfy the *incremental adjoint wave propagation equations*

$$\begin{aligned}
-\rho\frac{\partial\tilde{\boldsymbol{w}}}{\partial t} - \nabla\cdot(\mathsf{C}\tilde{\boldsymbol{D}}) &= \nabla\cdot(\tilde{\mathsf{C}}\boldsymbol{D}) - \mathcal{B}\tilde{\boldsymbol{v}} && \text{in } \Omega\times(0,T)\\
\mathsf{C}\frac{\partial\tilde{\boldsymbol{D}}}{\partial t} + \frac{1}{2}\mathsf{C}(\nabla\tilde{\boldsymbol{w}} + \nabla\tilde{\boldsymbol{w}}^T) &= \boldsymbol{0} && \text{in } \Omega\times(0,T)\\
\rho\tilde{\boldsymbol{w}} = \mathsf{C}\tilde{\boldsymbol{D}} &= \boldsymbol{0} && \text{in } \Omega\times\{t=T\}\\
\mathsf{C}\tilde{\boldsymbol{D}}\boldsymbol{n} &= -\tilde{\mathsf{C}}\boldsymbol{D}\boldsymbol{n} && \text{on } \Gamma\times(0,T)
\end{aligned}$$

Infinite dimensional Bayesian statistical inference

Linearized Bayesian solution: Low rank approximation

posterior covariance operator: A low rank approximation

$$\begin{aligned}
\mathcal{C} &= \left(A^*\Gamma^{-1}A + \mathcal{C}_0^{-1}\right)^{-1} \\
&= \mathcal{C}_0^{1/2}\left(\mathcal{C}_0^{1/2}\boldsymbol{A}^*\boldsymbol{\Gamma}^{-1}\boldsymbol{A}\,\mathcal{C}_0^{1/2} + \boldsymbol{I}\right)^{-1}\mathcal{C}_0^{1/2} \\
&\approx \mathcal{C}_0^{1/2}\left(\boldsymbol{V}_r\boldsymbol{\Lambda}_r\boldsymbol{V}_r^* + \boldsymbol{I}\right)^{-1}\mathcal{C}_0^{1/2} \\
&= \mathcal{C}_0 - \mathcal{C}_0^{1/2}\boldsymbol{V}_r\boldsymbol{D}_r\boldsymbol{V}_r^*\mathcal{C}_0^{1/2}
\end{aligned}$$

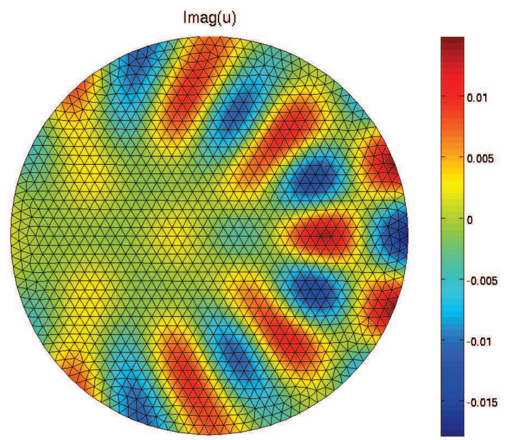
- Low rank approximation only involves incremetal forward and incremental adjoint solve
- Then use Sherman-Morrison-Woodbury
- Relative to the prior uncertainty, the posterior uncertainty is reduced when observations are made.

Linearized Bayesian solution: Why low rank approximation?

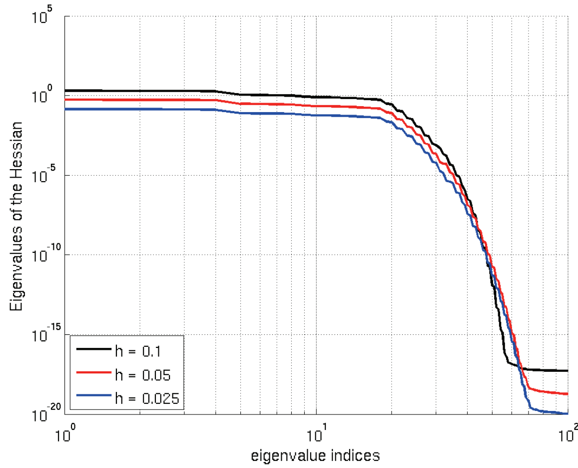
Compactness of the Hessian in inverse acoustic scattering

Theorem

Let $(1-n) \in C_0^{m,\alpha}$*, where* n *is the refractive index,* $m \in \mathbb{N} \cup \{0\}, \alpha \in (0,1)$*.*
The Hessian is a compact operator everywhere.



Coupled FEM-BEM method



Eigenvalues of Gauss-Newton Hessian

Details in:

- T. Bui-Thanh and O. Ghattas, *Analysis of the Hessian for inverse scattering problems. Part II: Inverse medium scattering of acoustic waves.* **Inverse Problems**, 28, 055002, 2012.
- T. Bui-Thanh and O. Ghattas, *Analysis of the Hessian for inverse scattering problems. Part I: Inverse shape scattering of acoustic waves.* **Inverse Problems 2012 Highlights Collection**, 28, 055001, 2012.

Infinite dimensional Bayesian statistical inference

Linearized Bayesian solution: Low rank approximation

posterior covariance operator: A low rank approximation

$$
\begin{aligned}
\mathcal{C} &= \left(A^*\Gamma^{-1}A + \mathcal{C}_0^{-1}\right)^{-1} \\
&= \mathcal{C}_0^{1/2}\left(\mathcal{C}_0^{1/2}\boldsymbol{A}^*\boldsymbol{\Gamma}^{-1}\boldsymbol{A}\,\mathcal{C}_0^{1/2} + \boldsymbol{I}\right)^{-1}\mathcal{C}_0^{1/2} \\
&\approx \mathcal{C}_0^{1/2}\left(\boldsymbol{V}_r\boldsymbol{\Lambda}_r\boldsymbol{V}_r^* + \boldsymbol{I}\right)^{-1}\mathcal{C}_0^{1/2} \\
&= \mathcal{C}_0 - \mathcal{C}_0^{1/2}\boldsymbol{V}_r\boldsymbol{D}_r\boldsymbol{V}_r^*\mathcal{C}_0^{1/2}
\end{aligned}
$$

- Low rank approximation only involves incremetal forward and incremental adjoint solve
- Then use Sherman-Morrison-Woodbury
- Relative to the prior uncertainty, the posterior uncertainty is reduced when observations are made.

3D high-resolution local and global full-waveform inversion

Wish list

Our choice

unified acoustic-elastic wave solver

first-order system formulation

3D high-resolution local and global full-waveform inversion

Wish list

unified acoustic-elastic wave solver

parallel (strong) scalability of wave solver

Our choice

first-order system formulation

Time domain + Explicit time stepping

3D high-resolution local and global full-waveform inversion

Wish list

unified acoustic-elastic wave solver

parallel (strong) scalability of wave solver

many wave lengths→low dispersion and dissipation

Our choice

first-order system formulation

Time domain + Explicit time stepping

high-order (spectral) elements

3D high-resolution local and global full-waveform inversion

Wish list

unified acoustic-elastic wave solver

parallel (strong) scalability of wave solver

many wave lengths→low dispersion and dissipation

resolve varying wave speeds

Our choice

first-order system formulation

Time domain + Explicit time stepping

high-order (spectral) elements

nonconforming meshing

3D high-resolution local and global full-waveform inversion

Wish list

unified acoustic-elastic wave solver

parallel (strong) scalability of wave solver

many wave lengths→low dispersion and dissipation

resolve varying wave speeds

Our choice

first-order system formulation

Time domain + Explicit time stepping

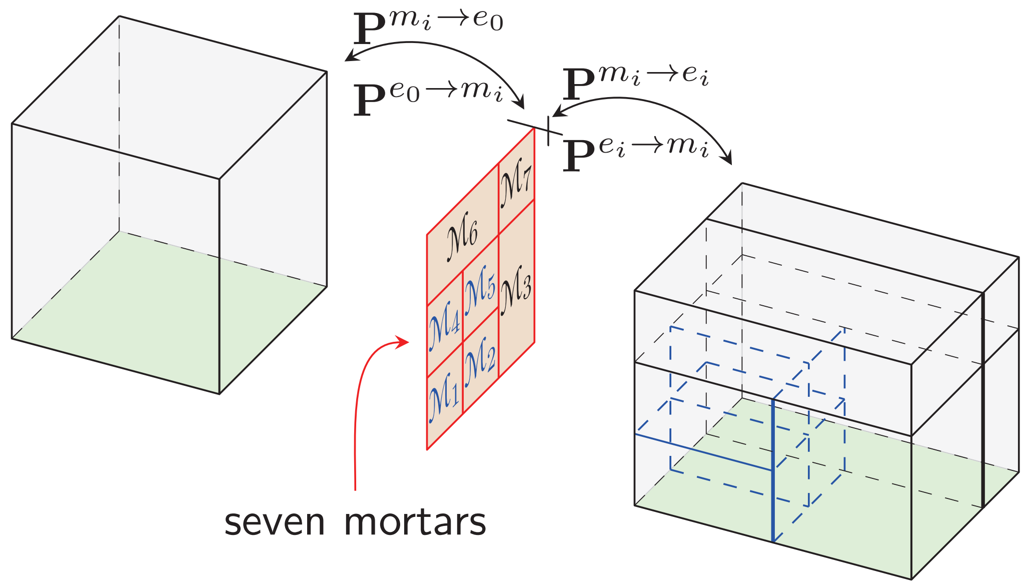
high-order (spectral) elements

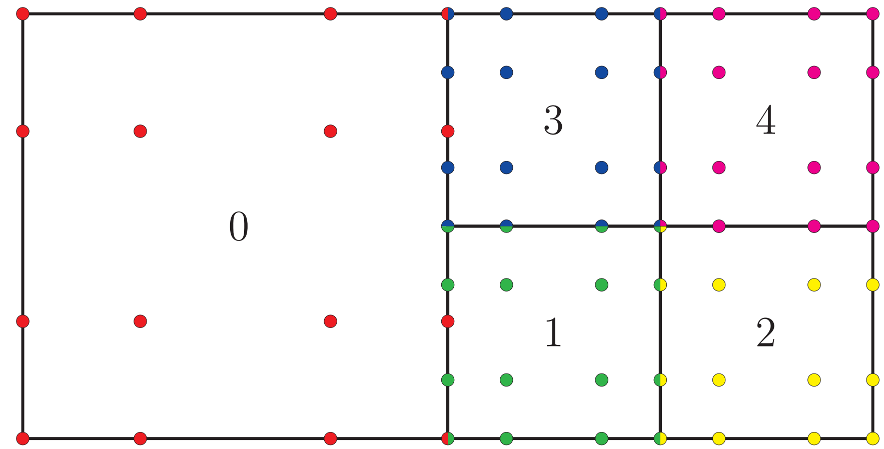
nonconforming meshing

Discontinuous Galerkin Spectral Element Method

hp non-conforming discontinuous Galerkin method

Elastic-Acoustic coupling as conservation laws





- nonconforming hexahedral elements using Kopriva's mortar approach for hyperbolic equations
  - theory for general hp-non-conforming dG
  - implementation for h-non-conforming dG with 2:1 balance
- tensor product Lagrange basis on the Legendre-Gauss-Lobatto (LGL) nodes
- LGL quadrature (diagonal mass matrix)
- time integration by classical 4-stage/RK4
- integrated parallel mesh generation/adaptivity

Convergence for non-conforming hp-discretization

Theorem

Assume $\mathfrak{q}^e \in \left[H^{s_e}\left(\mathsf{D}^e\right)\right]^d, s_e \geq 3/2$ *with* $d = 6$ *for electromagnetic case and* $d = 12$ *for elastic–acoustic case. In addition, suppose* $\mathfrak{q}_d(0) = \Pi\mathfrak{q}(0)$*, and the mesh is affine and non-conforming. Then, the discontinuous Galerkin spectral element solution* $\mathfrak{q}_d$ *converges to the exact solution* $\mathfrak{q}$*, i.e., there exists a constant* C *that depends only on the angle condition of* D^e*,* s*, and the material constants* μ *and* ε *(*λ *and* μ *for elastic–acoustic case) such that*

$$\left\|\mathfrak{q}\left(t\right) - \mathfrak{q}_d\left(t\right)\right\|_{\mathcal{D}^{N_{el}},d} \leq C \sum_e \frac{h_e^{\sigma_e}}{N_e^{s_e}} \left\|\mathfrak{q}\left(t\right)\right\|_{\left[H^{s_e}\left(\mathsf{D}^e\right)^d\right]} + C \sum_e t \frac{h_e^{\sigma_e - 1/2}}{N_e^{s_e - 1/2}} \max_{[0,t]} \left\|\mathfrak{q}\left(t\right)\right\|_{\left[H^{s_e}\left(\mathsf{D}^e\right)\right]^d},$$

with $h_e = diam\left(\mathsf{D}^e\right)$*,* $\sigma_e = \min\left\{p_e + 1, s_e\right\}$*, and* $\left\|\cdot\right\|_{H^s\left(\mathsf{D}^e\right)}$ *denoting the usual Sobolev norm*

Details in: T. Bui-Thanh and O. Ghattas, *Analysis of an hp-non-conforming discontinuous Galerkin spectral element method for wave propagations*, **SIAM Journal on Numerical Analysis**, 50(3), pp. 1801–1826, 2012.

Scalability of global seismic wave propagation on Jaguar

Strong scaling: 3rd order DG, 16,195,864 elements, 9.3 billion DOFs

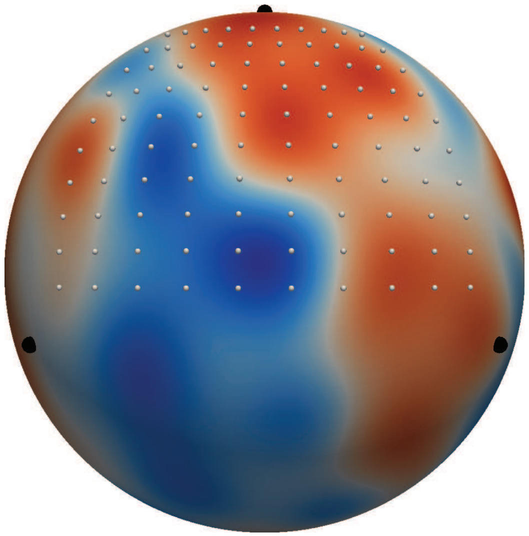
#cores	time [ms]	elem/core	efficiency [%]
1024	5423.86	15817	100.0
4096	1407.81	3955	96.3
8192	712.91	1978	95.1
16384	350.43	989	96.7
32768	211.86	495	80.0
65536	115.37	248	73.5
131072	57.27	124	74.0
262144	29.69	62	71.4

Strong scaling: 6th order DG, 170 million elements, 525 billion DOFs

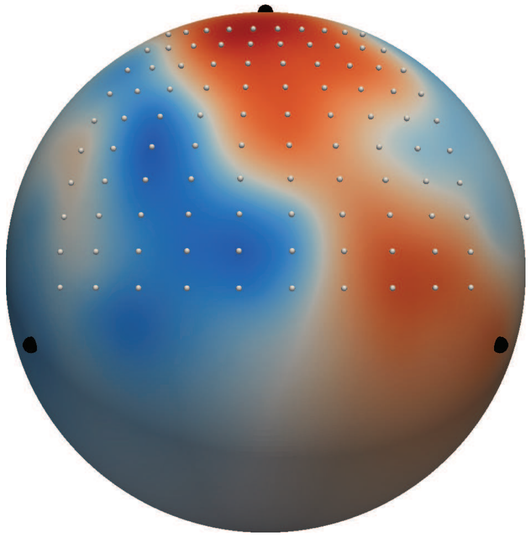
# cores	meshing time (s)	wave prop per step (s)	par eff wave	Tflops
32,640	6.32	12.76	1.00	25.6
65,280	6.78	6.30	1.01	52.2
130,560	17.76	3.12	1.02	105.5
223,752	$<$25	1.89	0.99	175.6

An example of global seismic inversion

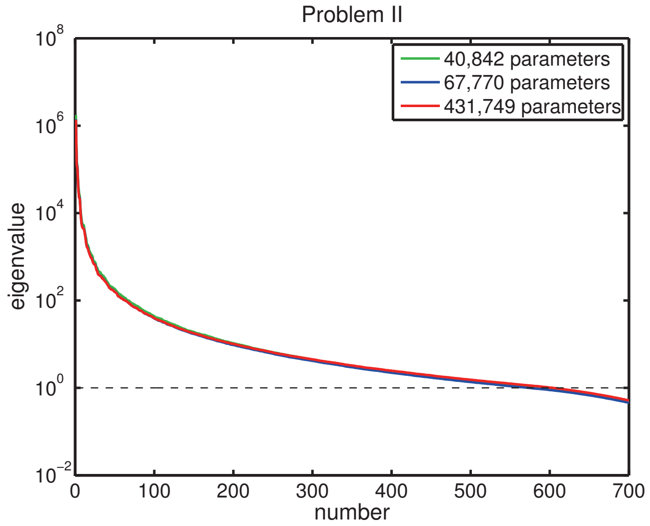
- inversion field: c_p in acoustic wave equation
- prior mean: PREM (radially symmetric model)
- "truth" model: S20RTS (Ritsema et al.), (laterally heterogeneous)
- Piecewise-trilinear on same mesh as forward/adjoint 3rd order dG fields
- dimensions: 1.07 million parameters, 630 million field unknowns
- Final time: $T = 1000s$ with 2400 time steps
- A single forward solve takes 1 minute on 64K Jaguar cores



"truth", sources (black)

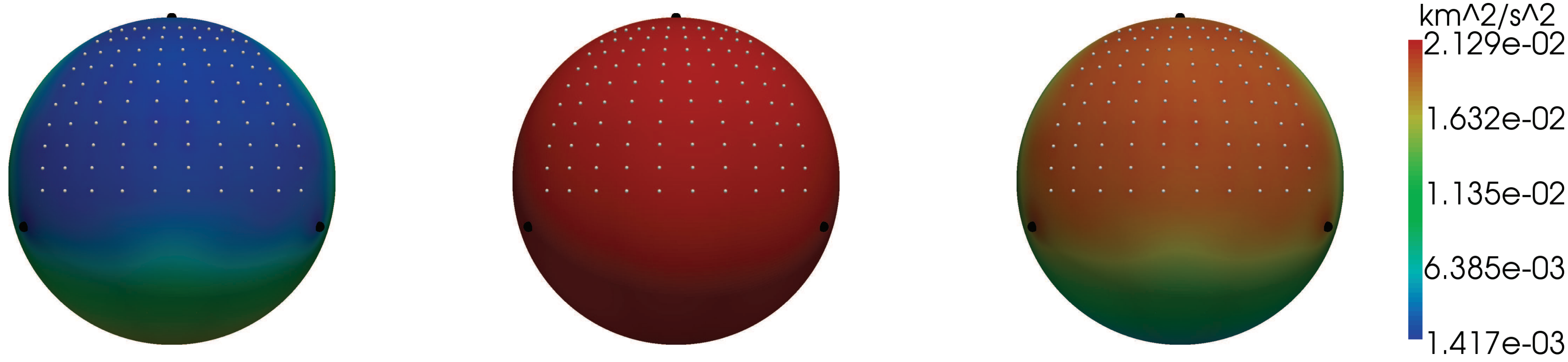


MAP, receivers (white)

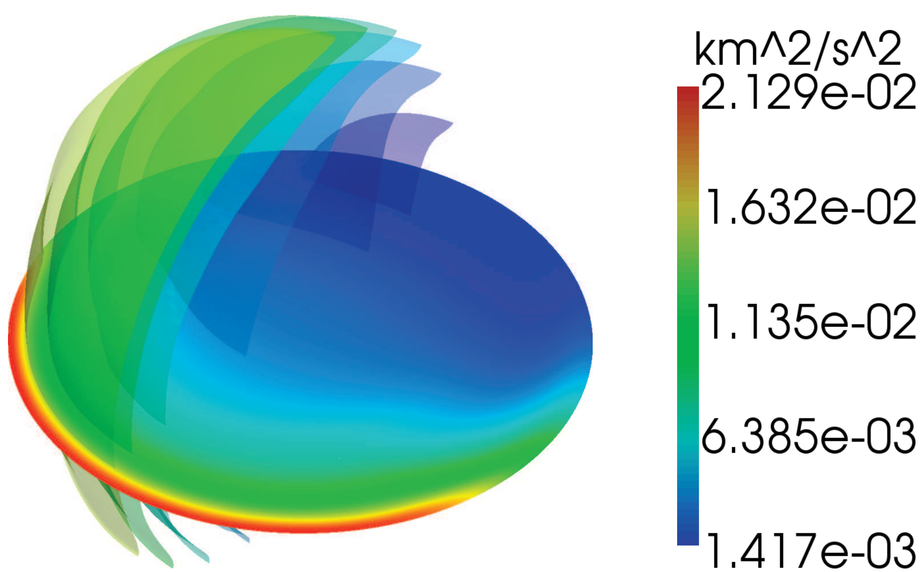


Hessian eigenvalues

Uncertainty quantification

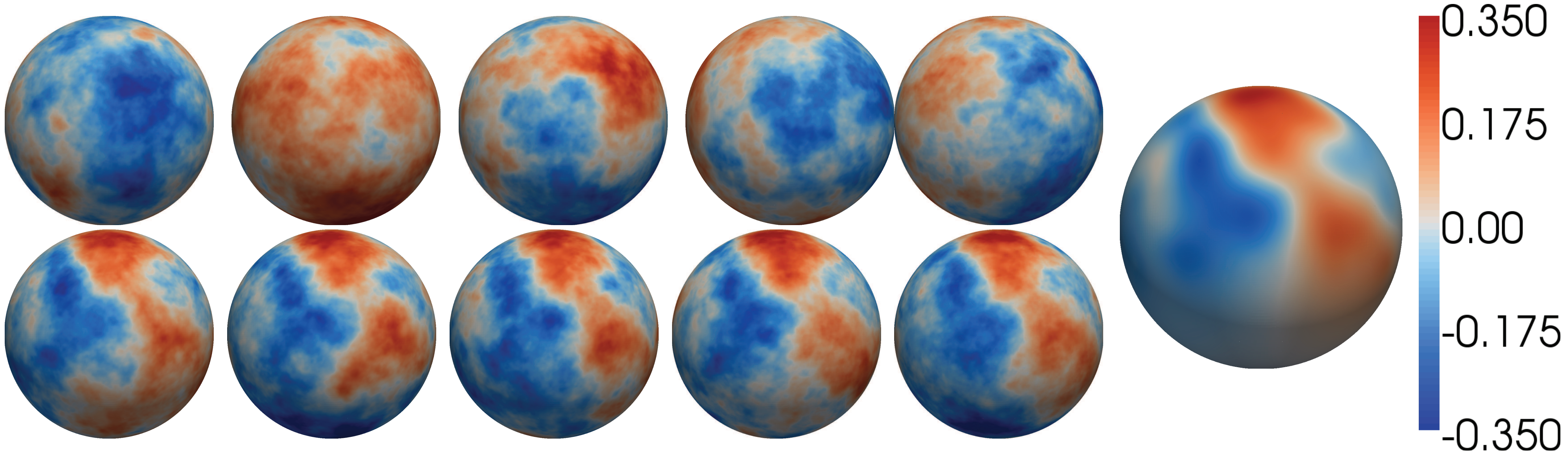


$$\mathfrak{C} \approx \mathfrak{C}_0 - \mathfrak{C}_0^{1/2} \boldsymbol{V}_r \boldsymbol{D}_r \boldsymbol{V}_r^* \mathfrak{C}_0^{1/2}$$



A slice through the equator and isosurfaces in the left hemisphere of variance reduction

Samples from prior and posterior distributions

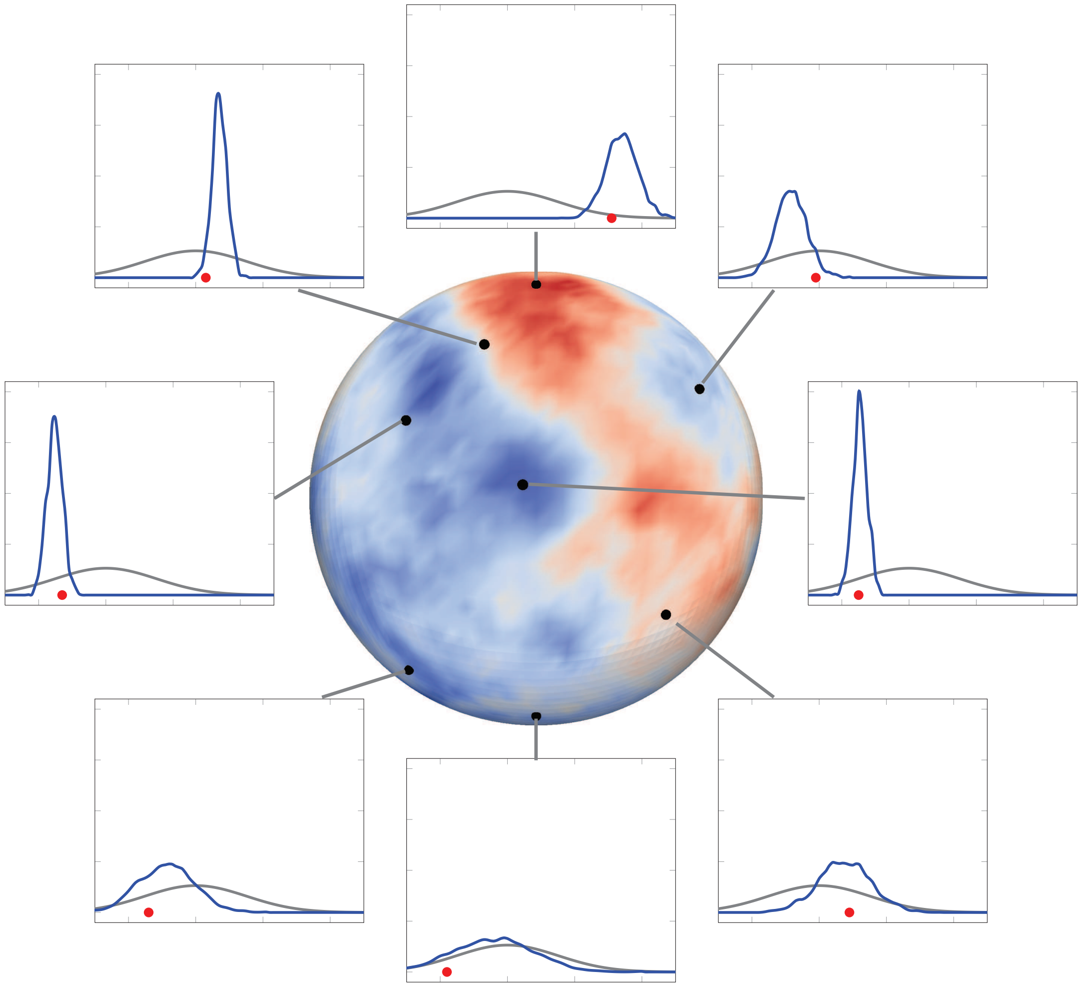


- **Top row**: samples from prior
- **Bottom row**: samples from posterior
- **Far right**: MAP estimate

Details in:

- Bui-Thanh, T., Burstedde, C., Ghattas, O., Martin, J., Stadler, G., and Wilcox, L.C.,*Extreme-scale UQ for Bayesian inverse problems governed by PDEs*, ACM/IEEE Supercomputing SC12, **Gordon Bell Prize Finalist**, 2012.
- Bui-Thanh, T., Ghattas, O., Martin, J., and Stadler, G., *A computational framework for infinite-dimensional Bayesian inverse problems. Part I: The linearized case*, **SIAM Journal on Scientific Computing**, 35(6), pp. A2494–A2523, 2013.

MCMC Simulation for Seismic inversion



- prior distribution
- posterior distribution
- posterior sample

- Use Gaussian approximation as proposal
- 15,587 samples, acceptance rate 0.28
- 96 hours on 2048 cores

Discretization of infinite dimensional Bayesian inversion

Error analysis and uncertainty quantification for 2D inverse shape acoustic scattering

- Shape $r = \exp(\mathbf{u})$, where $\mathbf{u} \in C^{s,\alpha}[0, 2\pi]$, $s \geq 2$ and $0 \leq \alpha \leq 1$
- Discretize μ_0 using Karhunen-Loève truncation with m terms
- Discretize the forward equation using n-th order Nyström scheme

Theorem

$$d_{\mathsf{Hellinger}}(\mu, \mu_{n,m}) \leq c\left(\frac{1}{(2n)^{s-1}} + \frac{\log m}{m^{s-1+\alpha}}\right),$$

$$\|\mathcal{E}_M\|_{L^2[0,2\pi]} \leq c\left(\frac{1}{(2n)^{s-1}} + \frac{\log m}{m^{s-1+\alpha}}\right),$$

$$\|\mathcal{E}_C\|_{L^2[0,2\pi]\otimes L^2[0,2\pi]} \leq c\left(\frac{1}{(2n)^{s-1}} + \frac{\log m}{m^{s-1+\alpha}}\right).$$

Details in: Bui-Thanh, T., and Ghattas, O., *An Analysis of Infinite Dimensional Bayesian Inverse Shape Acoustic Scattering and its Numerical Approximation*, **SIAM Journal on Uncertainty Quantification**, 2, pp. 203–222, 2014.

Outline

Current/Future Work

Forward algorithms/solvers at extreme scale

1. Hybridized discontinuous Galerkin method

Current/Future Work

Forward algorithms/solvers at extreme scale

1. Hybridized discontinuous Galerkin method

New MCMC algorithms for large-scale Bayesian inversions+big data

1. Riemannian manifold Hamiltonian MCMC
2. Randomized Maximum Likelihood
3. Finite element discretization of infinite dimensional MCMCs
4. UQ with big data

- Bui-Thanh, T., *From Godunov to A Unified Hybridized Discontinuous Galerkin Framework*, **Submitted**, 2014.
- Bui-Thanh, T., and Girolami, M., *Solving Large-scale PDE-Constrained Bayesian Inverse Problems With Riemann Manifold Hamiltonian Monte Carlo*, **Inverse Problems**, To Appear, 2014.
- Bui-Thanh, T., *On Finite Element Approximation of PDE-constrained Infinite Dimensional Bayesian Inverse Problems*, **Submitted**, 2014.
- Bui-Thanh, T., and Ghattas, O., *Randomized Maximum likelihood method for large-scale Bayesian Inversion*, **In Preparation**, 2014.

Conclusions: Reduce-Then-Sample

Computational Fluid dynamics: summary

- CFD is expensive for probabilistic purposes
- Reduced basis (Reduced-order) modeling seems to be useful
- Probabilistic analysis with reduced model is cost effective (orders of magnitude less time consuming), yet accurate

Computational Electromagnetics: summary

1. Statistical inversion via the Bayesian framework
2. Monte Carlo sampling the posterior in high dimensions is impractical
3. Hessian-based Piecewise Gaussian approximation to the posterior
4. Inverse solution comes with quantifiable uncertainty and more

Conclusions: Sample-Then-Reduce

Full wave form seismic inversion

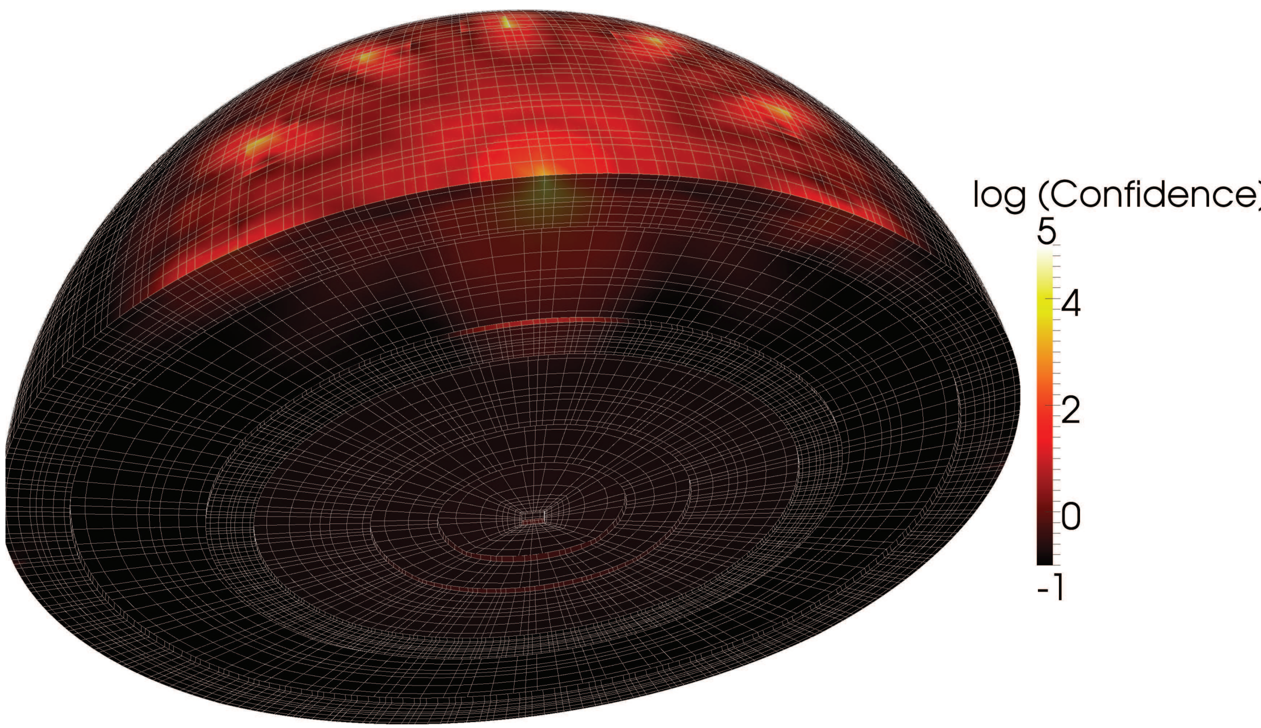
1. Given noise-corrupted measurements, want to image the earth interior
2. Infinite dimensional Bayesian inference
3. Explore the posterior measure: Derivation of $A^*\Gamma^{-1}A$
4. Compactness of $A^*\Gamma^{-1}A$
5. Scalable Discontinuous Galerkin for seismic waves: Scalability of A

Main results

- Able to solve statistical inverse problem with more than one million parameters with more than three orders of magnitude speedup
- Gaussian approximation seems to be good in this case
- Inverse solution comes with quantifiable uncertainty and more

Future work: Variance-driven adaptivity

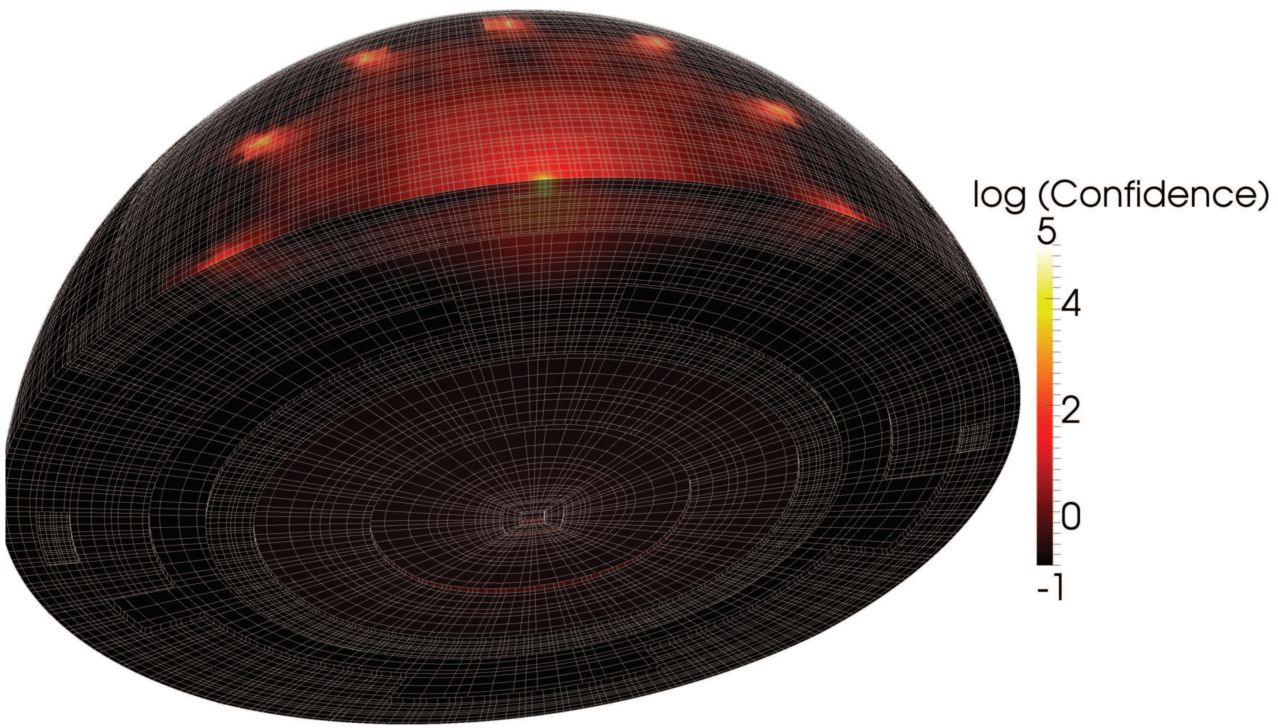
- adaptively refine parameter mesh based on posterior covariance (estimated by the inverse of the data-misfit Hessian)



- begin with coarse parameter discretization relative to wave field mesh
- locally refine parameter mesh based on variance information
- allows information contained in the data to drive the medium parametrization

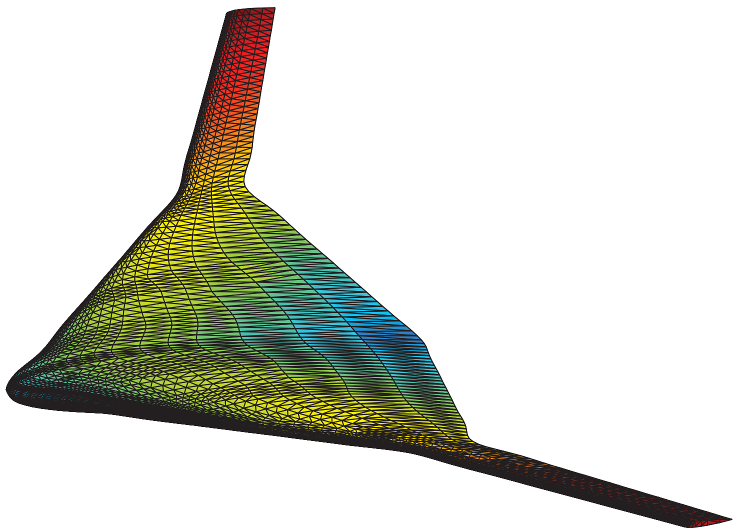
Future work: Variance-driven adaptivity

- adaptively refine parameter mesh based on posterior covariance (estimated by the inverse of the data-misfit Hessian)



- begin with coarse parameter discretization relative to wave field mesh
- locally refine parameter mesh based on variance information
- allows information contained in the data to drive the medium parametrization

Future work: Inverse Electromagnetic Scattering

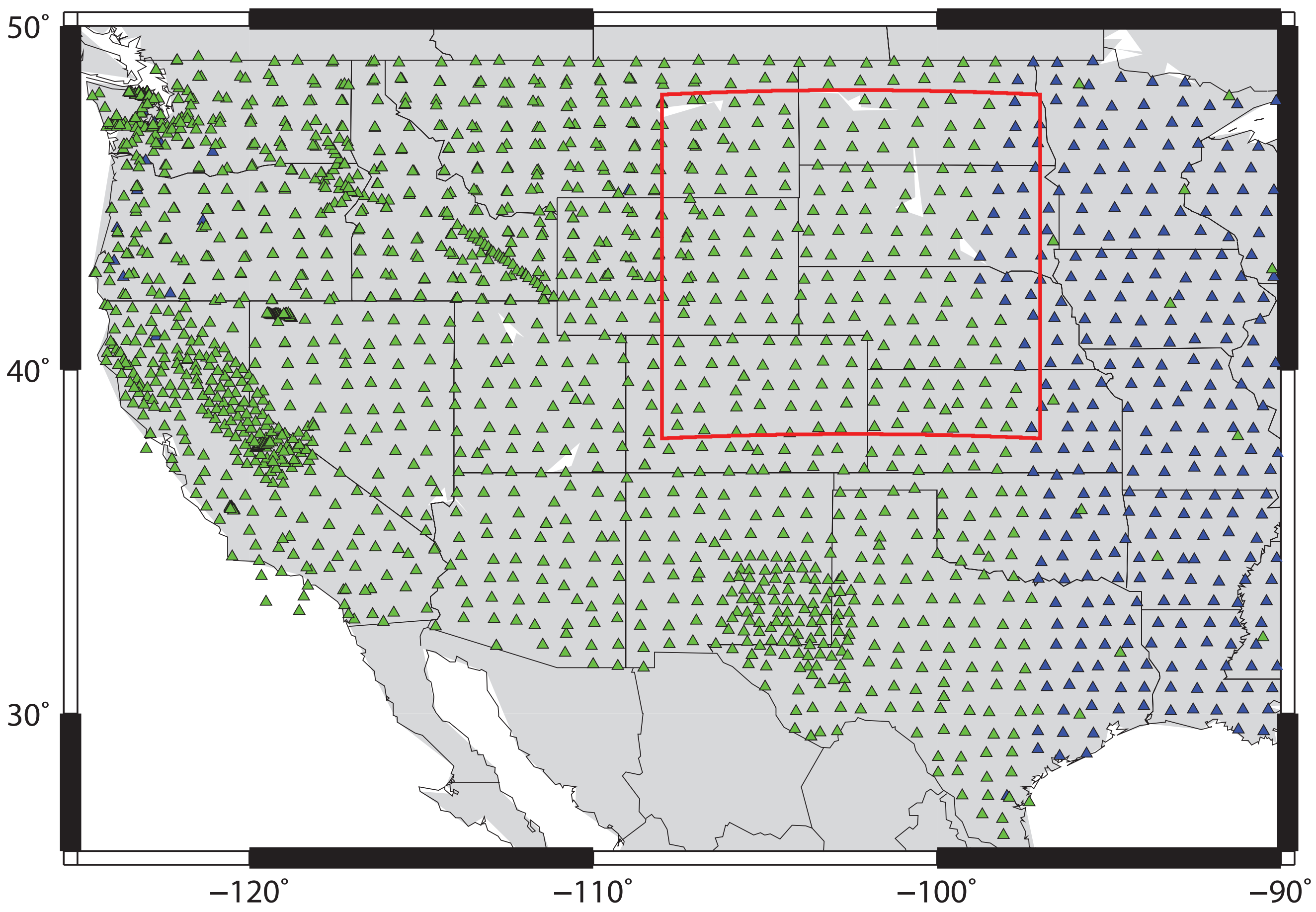


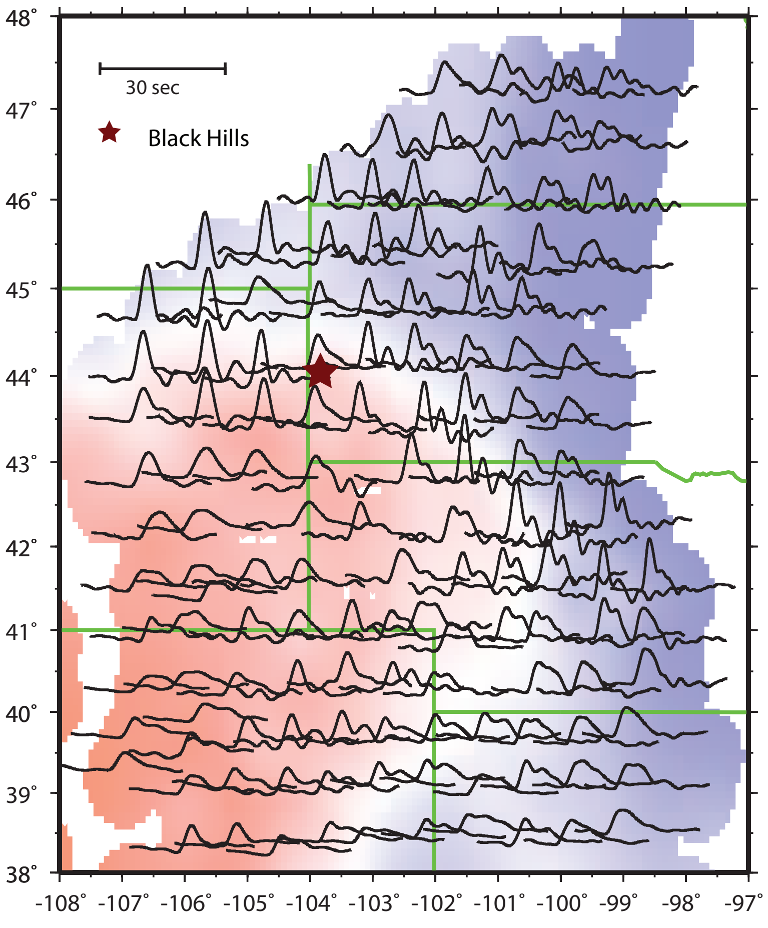
Blended wing body aircraft

Animated by Greg Abram

Details in: Bui-Thanh, T., Burstedde, C., and Ghattas, O., *A discontinuous Galerkin method for electromagnetic wave propagation on h-non-conforming adapted meshes*, In preparation, 2013.

Future work: Seismic inversion with real data





- Anisotropy
- Attenuation

Future work: Infinite dimensional MCMC on function spaces

- Infinite dimensional scaled stochastic Newton
- Infinite dimensional randomized maximum likelihood
- Infinite dimensional hybrid Monte Carlo
- Infinite dimensional Metropolis-adjusted Langevin method
- Infinite dimensional random walk

Future Research

Methodology

- Reduced basis (reduced-order modeling) methods
- Scalable response surface approach
- (Infinite dimensional) MCMC on function spaces
- Optimization-based discontinuous finite element method

Details of MCMC for seismic inversion

1. Use Gaussian approximation at MAP as a proposal for MCMC
2. Sampling performance for a coarser problem (with 78k parameters):
3. 15,587 MCMC samples (each requires 1 forward PDE solve) 4399 samples accepted (28%)
4. Integrated autocorrelation time of about 16–20) effective sample size of about 800
5. Total runtime of about 96 hours on 2048 cores

Infinite dimensional Bayesian statistical inference

Problem with the usual Bayes' formula

The usual Bayes' theorem

$$\pi_{\text{post}}\left(\mathbf{u}|\boldsymbol{y}^{\text{obs}}\right) \propto \pi_{\text{prior}}\left(\mathbf{u}\right) \pi_{\text{like}}\left(\boldsymbol{y}^{\text{obs}}|\mathbf{u}\right)$$

What are π's?

- They are probability density with respect to the Lebesgue's measure in $\mathbb{R}^n$
- There is no infinite dimensional Lebesgue's measure!

Infinite dimensional Bayesian statistical inference

Problem with the usual Bayes' formula

The usual Bayes' theorem

$$\pi_{\text{post}}\left(\mathbf{u} \middle| \boldsymbol{y}^{\text{obs}}\right) \propto \pi_{\text{prior}}\left(\mathbf{u}\right) \pi_{\text{like}}\left(\boldsymbol{y}^{\text{obs}} \middle| \mathbf{u}\right)$$

What are π's?

- They are probability density with respect to the Lebesgue's measure in $\mathbb{R}^n$
- There is no infinite dimensional Lebesgue's measure!

$$\cancel{\pi_{\text{post}}\left(\mathbf{u} \middle| \boldsymbol{y}^{\text{obs}}\right) \propto \pi_{\text{prior}}\left(\mathbf{u}\right) \pi_{\text{like}}\left(\boldsymbol{y}^{\text{obs}} \middle| \mathbf{u}\right)}$$

Infinite dimensional Bayesian statistical inference

Radon-Nikodym derivative

Define

- μ_0 the prior measure: $\frac{d\mu_0}{d\lambda} = \pi_{\text{prior}}$, where λ is the Lebesgue measure in $\mathbb{R}^n$
- $\mu^{\boldsymbol{y}^{\text{obs}}}$ the posterior measure conditional on $\boldsymbol{y}^{\text{obs}}$

Rewrite the usual Bayes' theorem using Radon-Nikodym derivative

$$\frac{d\mu^{\boldsymbol{y}^{\text{obs}}}}{d\mu_0}(\mathbf{u}) \propto \pi_{\text{like}}(\boldsymbol{y}^{\text{obs}}|\mathbf{u}) \stackrel{\text{def}}{=} \exp\Big(-\Phi\left(\boldsymbol{y}^{\text{obs}}, \mathbf{u}\right)\Big)$$

Infinite dimensional Bayesian statistical inference

Radon-Nikodym derivative

Define

- μ_0 the prior measure: $\frac{d\mu_0}{d\lambda} = \pi_{\text{prior}}$, where λ is the Lebesgue measure in $\mathbb{R}^n$
- $\mu^{\boldsymbol{y}^{\text{obs}}}$ the posterior measure conditional on $\boldsymbol{y}^{\text{obs}}$

Rewrite the usual Bayes' theorem using Radon-Nikodym derivative

$$\frac{d\mu^{\boldsymbol{y}^{\text{obs}}}}{d\mu_0}(\mathbf{u}) \propto \pi_{\text{like}}(\boldsymbol{y}^{\text{obs}}|\mathbf{u}) \stackrel{\text{def}}{=} \exp\left(-\Phi\left(\boldsymbol{y}^{\text{obs}}, \mathbf{u}\right)\right)$$

- This formulation is valid for infinite dimensional setting
- This formulation becomes the usual Bayes' theorem in finite dimensions

Others: Discontinuous Petrov-Galerkin (DPG) Method

Optimal convergence $p + 1$ as opposed to DG with $p + 1/2$

Details in:

- Roberts, N., Bui-Thanh, T., and Demkowicz, D., *The DPG Method for the Stokes Problem*, **Computers & Mathematics with Applications**, Submitted, 2012.
- Chan, J., Heuer, N., Bui-Thanh, T., and Demkowicz, D., *Robust DPG method for convection-dominated diffusion problems II: a natural in flow condition*, **Computers & Mathematics with Applications**, Accepted, 2012.
- Bui-Thanh, T., Demkowicz, L., and Ghattas, O., *A Unified Discontinuous Petrov-Galerkin Method and its Analysis for Friedrichs' Systems*, **SIAM Journal on Numerical Analysis**, Revised, 2012.
- Bui-Thanh, T., Demkowicz, L., and Ghattas, O., *Constructively Well-Posed Approximation Methods with Unity Inf-Sup and Continuity Constants for Partial Differential Equations*, **Mathematics of Computation**, To appear, 2012.

Gauss-Newton Hessian of data misfit with respect to c_p

$$D^2\mathcal{J}(c_p, c_s)(\delta c_p, \delta\tilde{c}_p, \delta\tilde{c}_s) := \sum_{\mathsf{D}^e}\left[\int_{\mathsf{D}^e}\int_0^T \tilde{V}_p^e \delta c_p + \int_{\partial\mathsf{D}^e}\int_0^T \tilde{S}_p^e \delta c_p^-\right] d\boldsymbol{x}dt$$

where:

$$\tilde{V}_p^e = 2\rho c_p \left(\boldsymbol{\nabla}\cdot\tilde{\boldsymbol{w}}\right)\mathrm{tr}(\boldsymbol{E})$$

$$\begin{aligned}\tilde{S}_p^e = & -k_0^2\rho^-\boldsymbol{n}\cdot[\![\boldsymbol{S}]\!]\boldsymbol{n}\cdot[\![\tilde{\boldsymbol{G}}]\!] \\ & + k_0^2\rho^-\left(\rho^+{c_p}^+\right)^2[\![\boldsymbol{v}]\!][\![\tilde{\boldsymbol{w}}]\!] \\ & + k_0^2\rho^-\rho^+{c_p}^+\left(\boldsymbol{n}\cdot[\![\boldsymbol{S}]\!][\![\tilde{\boldsymbol{w}}]\!] - [\![\boldsymbol{v}]\!]\boldsymbol{n}\cdot[\![\tilde{\boldsymbol{G}}]\!]\right) \\ & + 2k_0\rho^-{c_p}^-\,\mathrm{tr}(\boldsymbol{E}^-)\left(\boldsymbol{n}\cdot[\![\tilde{\boldsymbol{G}}]\!] - \rho^+{c_p}^+[\![\tilde{\boldsymbol{w}}]\!]\right)\end{aligned}$$

$$k_0 = \frac{1}{\rho^- c_p^- + \rho^+ c_p^+}$$

Gauss-Newton Hessian of data misfit with respect to c_s

$$D^2 \mathcal{J}(c_p, c_s)(\delta cs, \delta\tilde{c_p}, \delta\tilde{c_s}) := \sum_{\mathsf{D}^e} \left[\int_{\mathsf{D}^e} \int_0^T \tilde{V_s^e} \delta cs + \int_{\partial \mathsf{D}^e} \int_0^T \tilde{S_s^e} \delta cs^- \right] d\boldsymbol{x} dt$$

where:

$$\tilde{V_s^e} = 4\rho c_s \left(\boldsymbol{E} - \mathrm{tr}(\boldsymbol{E}) \boldsymbol{I} \right) : \frac{1}{2} \left(\nabla \tilde{\boldsymbol{w}} + \nabla \tilde{\boldsymbol{w}}^T \right)$$

$$\begin{aligned}
\tilde{S_s^e} = & - k_1^2 \rho^- \left(\boldsymbol{n} \times (\boldsymbol{n} \times [\![\boldsymbol{S}]\!]) \right) \cdot \left(\boldsymbol{n} \times \left(\boldsymbol{n} \times [\![\tilde{\boldsymbol{G}}]\!] \right) \right) \\
& + k_1^2 \rho^- \left(\rho^+ {c_s}^+ \right)^2 \left(\boldsymbol{n} \times (\boldsymbol{n} \times [\boldsymbol{v}]) \right) \cdot \left(\boldsymbol{n} \times (\boldsymbol{n} \times [\tilde{\boldsymbol{w}}]) \right) \\
& + k_1^2 \rho^- \rho^+ {c_s}^+ \left(\boldsymbol{n} \times (\boldsymbol{n} \times [\![\boldsymbol{S}]\!]) \right) \cdot \left(\boldsymbol{n} \times (\boldsymbol{n} \times [\tilde{\boldsymbol{w}}]) \right) \\
& - k_1^2 \rho^- \rho^+ {c_s}^+ \left(\boldsymbol{n} \times (\boldsymbol{n} \times [\boldsymbol{v}]) \right) \cdot \left(\boldsymbol{n} \times \left(\boldsymbol{n} \times [\![\tilde{\boldsymbol{G}}]\!] \right) \right) \\
& + 4 k_0 \rho^- {c_s}^- \left(\boldsymbol{n} \cdot \boldsymbol{E}^- \boldsymbol{n} - \mathrm{tr}(\boldsymbol{E}^-) \right) \left(\boldsymbol{n} \cdot [\![\tilde{\boldsymbol{G}}]\!] - \rho^+ {c_p}^+ [\![\tilde{\boldsymbol{w}}]\!] \right) \\
& + 4 k_1 \rho^- {c_s}^- \left(\boldsymbol{n} \times (\boldsymbol{n} \times \boldsymbol{E}^- \boldsymbol{n}) \right) \cdot \left(\boldsymbol{n} \times \left(\boldsymbol{n} \times \left([\![\tilde{\boldsymbol{G}}]\!] - \rho^+ {c_s}^+ [\tilde{\boldsymbol{w}}] \right) \right) \right)
\end{aligned}$$

State-of-the-art Optimization Technique

Subspace Trust Region Interior Reflective Inexact Newton-CG

1. Trust region (deals with ill-conditioning)
2. Interior reflective (deals with bound constraints)
3. Inexact Newton-CG (converges quadratically, prevents oversolving)
4. Uses adjoint method to compute the gradient
5. Computes Hessian-vector product on-the-fly using one forward-like and one adjoint-like solves

$\longrightarrow$ Efficient optimization solver

State-of-the-art Optimization Technique

Subspace Trust Region Interior Reflective Inexact Newton-CG

1. Trust region (deals with ill-conditioning)
2. Interior reflective (deals with bound constraints)
3. Inexact Newton-CG (converges quadratically, prevents oversolving)
4. Uses adjoint method to compute the gradient
5. Computes Hessian-vector product on-the-fly using one forward-like and one adjoint-like solves

$\longrightarrow$ Efficient optimization solver

The number of Newton steps is numerically observed $\mathcal{O}(d)$.

d: the number of parameters

GPU implementation: AMR on CPU, wave prop on GPU

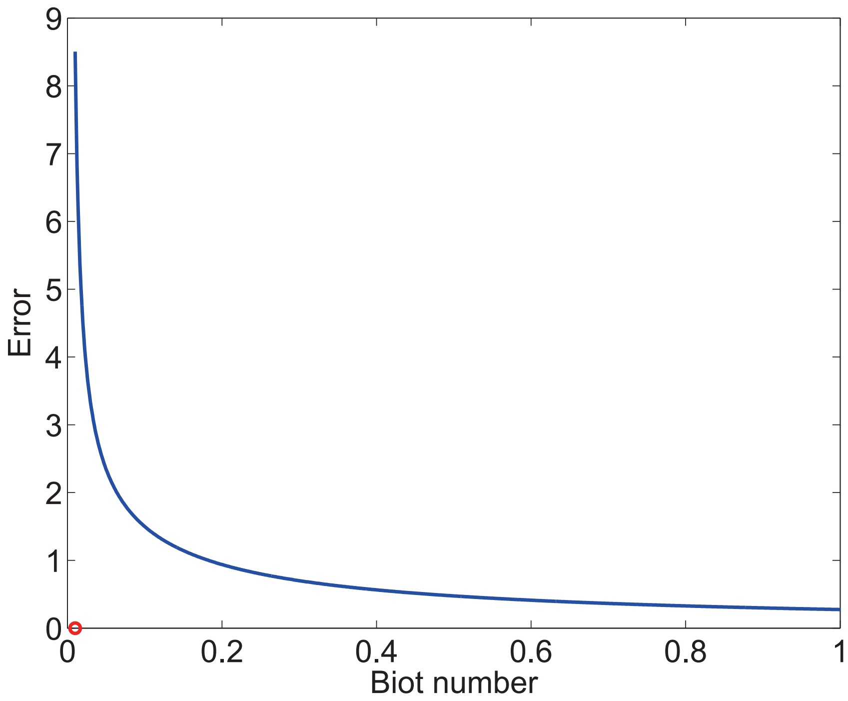
Excellent weak scalability on TACC's GPU cluster (collaboration with T. Warburton, Rice)

#GPUs	#elem	mesh (s)	transf (s)	wave prop	par eff wave	Tflops (s.p.)
8	224048	9.40	13.0	29.95	1.000	0.63
64	1778776	9.37	21.3	29.88	1.000	5.07
256	6302960	10.6	19.1	30.03	0.997	20.3
478	12270656	11.5	16.2	29.89	1.002	37.9

- Up to 12.3 million 7-th order elements (67 billion unknowns)
- *transf* indicates the time to transfer the mesh and other initial data from CPU to GPU memory
- *wave prop* is the runtime in μsec per time step per average number of elements per GPU
- wallclock time about 1 second per time step (meshing and transfer time are completely negligible for realistic simulations)
- *Longhorn* = 512 NVIDIA FX 5800 GPUs each with 4GB graphics memory and 512 Intel Nehalem quad core processors connected by QDR InfiniBand interconnect

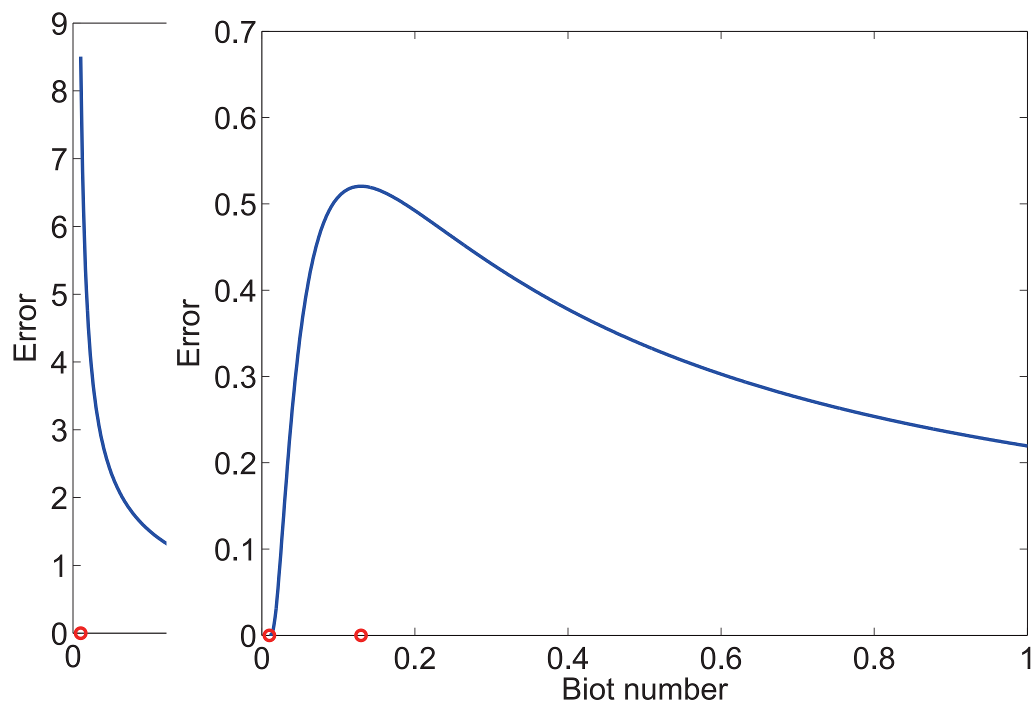
The Model-Constrained Adaptive Sampling Algorithm

1. Given a reduced basis Φ and initial guess $\mathbf{u}^0$. Find $\mathbf{u}^* = \arg\max_{\mathbf{u}} \|y(\mathbf{u}) - y_r(\mathbf{u})\|_2^2$
2. If $\|y(\mathbf{u}^*) - y_r(\mathbf{u}^*)\|_2^2 \leq \epsilon$, then terminate the algorithm. If not, go to the next step.
3. With $\mathbf{u} = \mathbf{u}^*$, solve the full system to compute the state solutions $x(\mathbf{u}^*)$, which is then used to update the basis Φ. Go to step 1.



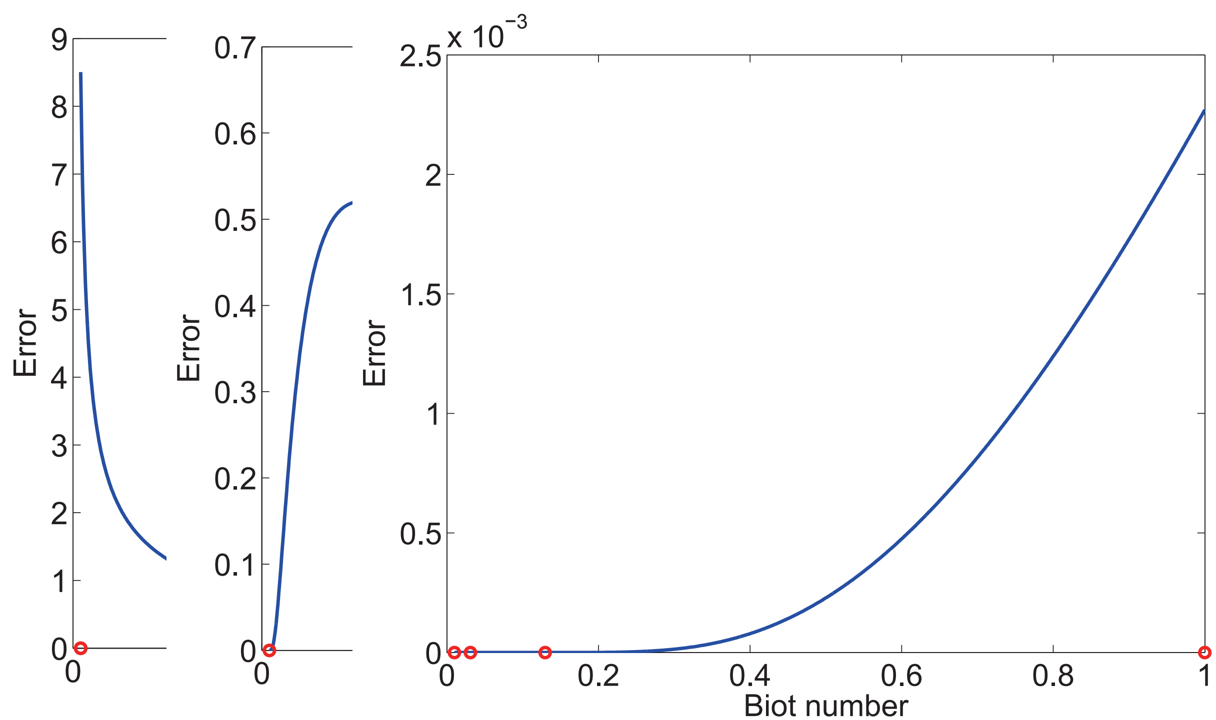
The Model-Constrained Adaptive Sampling Algorithm

1. Given a reduced basis Φ and initial guess $\mathbf{u}^0$. Find $\mathbf{u}^* = \arg\max_{\mathbf{u}} \|y(\mathbf{u}) - y_r(\mathbf{u})\|_2^2$
2. If $\|y(\mathbf{u}^*) - y_r(\mathbf{u}^*)\|_2^2 \leq \epsilon$, then terminate the algorithm. If not, go to the next step.
3. With $\mathbf{u} = \mathbf{u}^*$, solve the full system to compute the state solutions $x(\mathbf{u}^*)$, which is then used to update the basis Φ. Go to step 1.



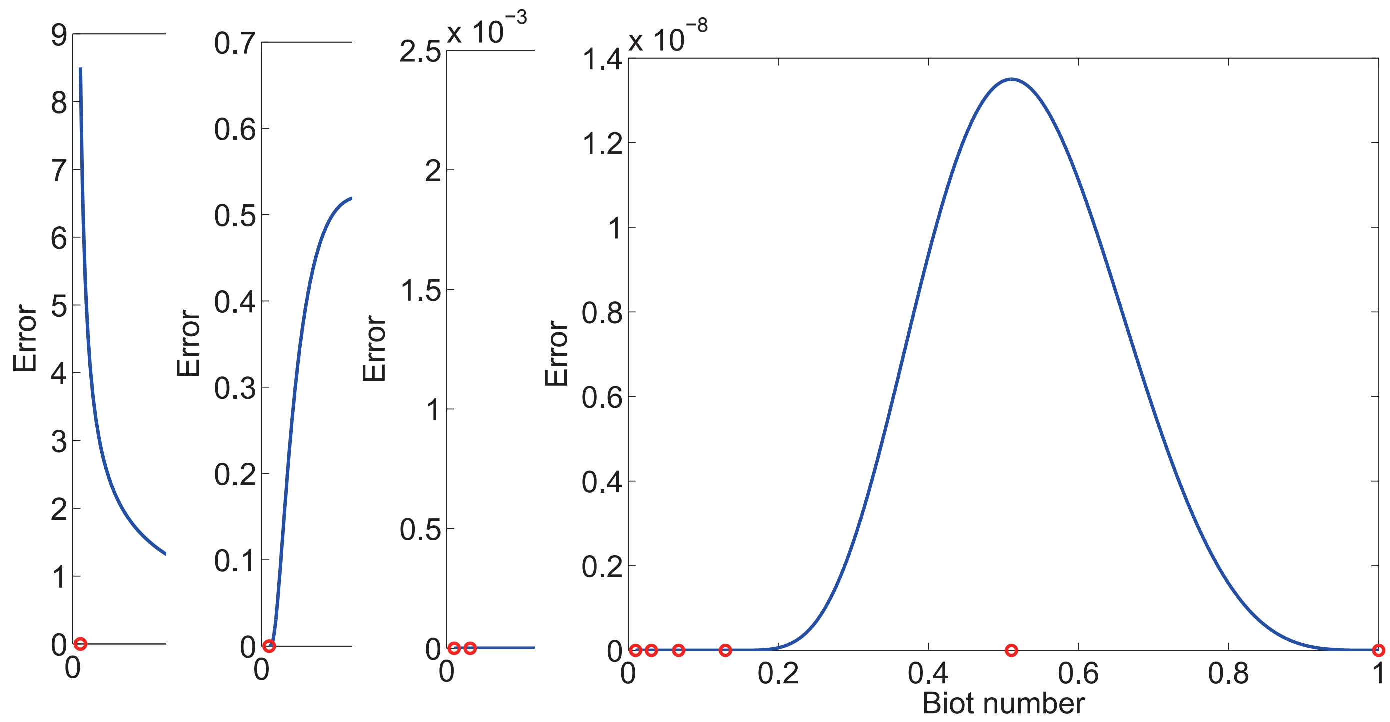
The Model-Constrained Adaptive Sampling Algorithm

1. Given a reduced basis Φ and initial guess $\mathbf{u}^0$. Find $\mathbf{u}^* = \arg\max_{\mathbf{u}} \|y(\mathbf{u}) - y_r(\mathbf{u})\|_2^2$
2. If $\|y(\mathbf{u}^*) - y_r(\mathbf{u}^*)\|_2^2 \leq \epsilon$, then terminate the algorithm. If not, go to the next step.
3. With $\mathbf{u} = \mathbf{u}^*$, solve the full system to compute the state solutions $x(\mathbf{u}^*)$, which is then used to update the basis Φ. Go to step 1.



The Model-Constrained Adaptive Sampling Algorithm

1. Given a reduced basis Φ and initial guess $\mathbf{u}^0$. Find $\mathbf{u}^* = \arg\max_{\mathbf{u}} \|y(\mathbf{u}) - y_r(\mathbf{u})\|_2^2$
2. If $\|y(\mathbf{u}^*) - y_r(\mathbf{u}^*)\|_2^2 \leq \epsilon$, then terminate the algorithm. If not, go to the next step.
3. With $\mathbf{u} = \mathbf{u}^*$, solve the full system to compute the state solutions $x(\mathbf{u}^*)$, which is then used to update the basis Φ. Go to step 1.



The Model-Constrained Adaptive Sampling Algorithm

1. Given a reduced basis Φ and initial guess $\mathbf{u}^0$. Find $\mathbf{u}^* = \arg\max_{\mathbf{u}} \|y(\mathbf{u}) - y_r(\mathbf{u})\|_2^2$
2. If $\|y(\mathbf{u}^*) - y_r(\mathbf{u}^*)\|_2^2 \leq \epsilon$, then terminate the algorithm. If not, go to the next step.
3. With $\mathbf{u} = \mathbf{u}^*$, solve the full system to compute the state solutions $x(\mathbf{u}^*)$, which is then used to update the basis Φ. Go to step 1.